Do less informative college admission exams reduce earnings inequality? Evidence from Colombia

Evan Riehl*
Cornell University

January 2023

Abstract. This paper asks whether reducing the informativeness of college admission exams reduces inequality in post-college earnings. I examine a 2000 reform of the Colombian college admission exam that sought to reduce bias in scores. The reform reduced test score gaps between high- and low-income students by nearly 50 percent in some subjects, but it also decreased the exam's predictive power for abilities that matter in college. I find that the reform caused students to attend colleges where they were more likely to drop out, which reduced earnings for *both* high- and low-income students.

* Department of Economics, Cornell University, 266 Ives Hall, Ithaca, NY 14853 (e-mail: eriehl@cornell.edu).
Previous versions of this paper were titled "Assortative Matching and Complementarity in College Markets" and "Fairness in College Admission Exams: From Test Score Gaps to Earnings Inequality." For useful comments I thank Peter Arcidiacono, Peter Bergman, Serena Canaan, David Deming, Nicolás De Roux, Christian Dustmann, Christopher Hansman, Adam Kapor, Lars Lefgren, Thomas Lemieux, Michael Lovenheim, W. Bentley MacLeod, Costas Meghir, Luis Omar Herrera Prada, Jonah Rockoff, Miikka Rokkanen, Juan E. Saavedra, Judith Scott-Clayton, Mallika Thomas, Miguel Urquiola, and Eric Verhoogen. All errors are my own.

Do standardized college admission exams reduce or exacerbate inequality? The debate over admission exams usually centers on how they affect access to selective colleges. In the U.S., the SAT exam was created with the goal of reducing the role of family wealth in admissions (Lemann, 2000), and yet there is a growing perception that such tests are biased against low-income and minority students. This backlash has led many colleges to reduce or eliminate the role of standardized exams in admissions. For example, the University of Chicago switched to test-optional admissions in 2018, and, in 2021, the University of California announced that it would no longer consider SAT or ACT scores at all.

This paper takes a different approach to this debate by asking how college admission exams affect inequality in labor market outcomes. I examine the case of the national college admission exam in the country of Colombia, which was redesigned in the year 2000 to address similar concerns about bias in scores. I show that this exam overhaul reduced test score gaps between high- and low-income students by nearly 50 percent in some subjects, but it did so primarily by becoming a worse measure of abilities that are important for college success. As a result, the reform caused students to attend colleges where they were more likely to drop out, which reduced labor market earnings for *both* high- and low-income students.

It is hard to know *ex ante* how the design of an admission exam affects earnings outcomes because there are many dimensions of heterogeneity in the returns to selective colleges. Some research finds that low-income students have large returns to attending selective colleges (e.g., Dale and Krueger, 2002; Chetty et al., 2020), which suggests that reforms that reduce socioeconomic gaps in admission exam scores may benefit disadvantaged students. But other work finds that students benefit from attending colleges with peers who have similar levels of academic preparation (Arcidiacono et al., 2016; Dillon and Smith, 2020). In this case, it is important to understand how admission reforms affect the measurement of student ability. Exams that are poor measures of ability can reduce the quality of student/college matches, leading to worse outcomes for all students.

My empirical analysis exploits a reform of the Colombian national college admission test, known as the ICFES exam. In 2000, the ICFES exam underwent a major overhaul with the goal of reducing socioeconomic bias in scores. The new exam was designed to test "competencies" rather than "content," and was similar in spirit to the new version of the U.S. SAT exam that was introduced in 2016. I use individual-level administrative data to measure the effects of this reform on students' test scores, college selectivity, and earnings outcomes measured more than a decade later.

I first show that the ICFES reform dramatically reduced socioeconomic test score gaps, but at the cost of becoming a worse predictor of students' college outcomes. For example, the test score gap between students in the top and bottom income quartiles fell by 50 percent on the math and physics components of the exam. But the exam's predictive power for students' college outcomes—including first-year GPA and the likelihood of graduating—declined by a similar magnitude. This suggests that test score gaps declined primarily because randomness became a more important determinant of scores, consistent with several features of the new exam that reduced its information content.

I use geographic variation in the stakes of the ICFES exam to identify the impacts of the reform on student. In Colombia, students typically attend college close to home, and the most selective colleges are large public flagship universities. At most public universities, admission is determined solely by scores on the ICFES exam, but some flagship schools use their own admission exams. These different admission methods created geographic variation in the stakes of the ICFES reform. I define students as "treated" by the reform if they lived near public universities that used ICFES scores in admissions during the pre-reform period. I define "control" students as those who lived near schools with other admission criteria. My empirical specification is a difference-in-differences model that estimates changes in outcomes across exam cohorts and treated/control areas.

My main finding is that the ICFES reform reduced graduation rates and earnings for *both* high- and low-income students. In treated areas, low-income students attended more selective colleges after the reform, and high-income students were displaced to less-selective schools. But both high- and low-income students experienced decreases in their post-college earnings (measured relative to earlier cohorts). Overall, earnings fell by 1–2 percent measured across *all* college students in treated areas, and these negative effects persisted up to 16 years after the ICFES exam. Graduation rates also declined for both groups. This suggests that the reform caused some students to drop out of college, which harmed them in the labor market.

Lastly, I present evidence that the ICFES reform reduced the quality of student/college matches in programs where the old exam was more predictive of student success. The decline in the ICFES exam's predictive power was most pronounced in its math and physics components. I show that the negative graduation and earnings effects were concentrated in quantitative fields like STEM and business, where pre-reform math and physics scores were stronger predictors of degree completion. This suggests that the reform caused some students to enroll in programs were they were academically underprepared, which can explain why graduation rates and earnings fell even for low-income students.

My paper contributes to a large literature on the returns to college selectivity.[1] By the nature of their empirical strategies, most papers in this literature estimate returns for students who are qualified for admission to a selective college. For example, many papers use regression discontinuity designs to estimate returns for students on the margin of admission

[1] See, for example, Saavedra (2009); Hastings et al. (2013); Zimmerman (2014); Andrews et al. (2016); Canaan and Mouganie (2018); Hoxby (2018); Zimmerman (2019); Anelli (2020); Sekhri (2020); Smith et al. (2020); Michelman et al. (2022); Ng and Riehl (2022).

(Hoekstra, 2009; Kirkebøen et al., 2016). Other work uses the "Dale and Krueger (2002) identification strategy," which compares students who were admitted to the same colleges but made different enrollment choices (Chetty et al., 2020; Mountjoy and Hickman, 2020). But the earnings effects of an admission exam reform depend precisely on students who were *not* qualified under the old exam. Further, large-scale reforms can alter the composition of a college's student body, which may change students' returns if peer effects are important (Arcidiacono and Vigdor, 2010; Machado et al., 2022). This can explain why many papers find that disadvantaged students have positive returns to selective colleges, and yet the ICFES reform reduced earnings for the average low-income student.

My paper also relates to research on "mismatch" in college enrollment (Arcidiacono and Lovenheim, 2016). This work finds that students have higher graduation rates and earnings when their own academic preparation is close to that of their classmates (Arcidiacono, 2004; Arcidiacono et al., 2016; Dillon and Smith, 2020).[2] Much attention is paid to a special case of mismatch, which hypothesizes that underprepared students have *negative* returns to selective colleges. But mismatch can arise more broadly even if all students have positive returns. If college selectivity and ability are complementary, it is important to have good measures of ability so that the students who benefit the most from attending selective colleges are the ones who gain admission. This can explain why my results differ from those in Black et al. (2023), who show that disadvantaged students benefited from Texas' Top 10 Percent policy. The authors find that students who were "pulled in" to top colleges by the policy were *more* prepared academically than the students they displaced, while the opposite was true in the case of the ICFES reform.

Lastly, my paper relates to work on the design of college admission exams (Rothstein, 2004; Bettinger et al., 2013; Hoxby and Turner, 2013; Bulman, 2015; Goodman, 2016; Goodman et al., 2020).[3] This work shows how exams affect college access, while my paper explores their labor market implications. My findings inform broader debates about the use of placement exams or GPA thresholds for controlling access to college programs (Scott-Clayton, 2012; Bergman et al., 2021; Bleemer and Mehta, 2021), and on the desirability of K–12 "exam schools" that use of admission tests to select students (Angrist et al., 2019).

The paper proceeds as follows. Section 1 discusses the mechanisms through which the design of a college admission exam can affect earnings outcomes. Section 2 describes the ICFES reform and its effects on test score gaps and predictive power. Section 3 describes my

[2] Other work exploits affirmative actions bans but finds no direct evidence of mismatch (Cortes, 2010; Backes, 2012; Hinrichs, 2012, 2014). A related literature examines the causes of "overmatch" or "undermatch" in college admissions (Hoxby and Avery, 2013; Smith et al., 2013; Dillon and Smith, 2017).

[3] Other matching mechanisms in education include centralized assignment (Abdulkadiroğlu et al., 2005), affirmative action (Durlauf, 2008; Bertrand et al., 2010; Bagde et al., 2016), and percent plans (Long, 2004; Kain et al., 2005; Niu and Tienda, 2010; Cullen et al., 2013; Daugherty et al., 2014; Kapor, 2015).

identification strategy, and Section 4 presents my main results on the reform's impacts on longer-run outcomes. Section 5 presents evidence on mechanisms, and Section 6 concludes.

1. Conceptual framework

This paper considers the impacts of an admission exam reform that reduces test score gaps between students from high and low socioeconomic backgrounds. Such a reform will tend to expand access to selective college programs for students with low socioeconomic status (SES), and it will tend to restrict the college choices of high SES students. This section discusses the mechanisms through which this change in college access can affect mean earnings measured across all students, as well as earnings for low SES students in particular. Appendix B describes these mechanisms more formally in a multinomial college choice framework.

The earnings effects of exam reforms depend on students' *returns to a preferred college program*. A student's exam score affects the set of programs that they can gain admission to, and thus they may be able to attend a more-preferred program if they would score higher on one version of the exam. The relevant return is the difference between a student's potential earnings at a more- and less-preferred program, where the programs that the student can attend are defined by their potential scores on the pre- and post-reform exams. Although many students would prefer programs that give them higher earnings, the return to a preferred program can be negative if students have incorrect expectations (Arcidiacono et al., 2011) or if they value non-pecuniary program characteristics (e.g., enjoyment of the material).

All else equal, an exam reform will increase mean earnings if the students who are "shifted in" to a preferred program by the reform have larger returns to those programs than those who are "shifted out" of their preferred programs. Conversely, mean earnings decrease if the students in the "shifted out" group have larger returns in aggregate. This intuition is similar to that in the conceptual frameworks in Black et al. (2023) and Dalla Zuanna et al. (2022).

Thus for a reform that reduces SES test score gaps, one important consideration is whether high or low SES students have larger returns to selective colleges. Some papers find that these returns are larger for low SES students (Dale and Krueger, 2002; Bleemer, 2022), possibly because they benefit more from selective schools' financial resources (Deming and Walters, 2017) or because they tend to have lower-ranked schools as their fallback option (Hoxby and Avery, 2013; Angrist et al., 2022). On the other hand, some work finds that high SES students have larger returns to elite colleges because they benefit more from peer networks (Zimmerman, 2019; Michelman et al., 2022).

In addition to heterogeneity by SES, the earnings impacts also depend on how the exam redesign affects the measurement of ability. Exam designers often aim to reduce the importance of "test prep," which may reduce SES score gaps while also improving the exam's measurement of abilities that are important to succeed in college. But SES gaps can also decline because randomness becomes relatively more important. For example, if the exam becomes too hard or too easy, much of the variation in scores is driven by guessing. This will also reduce SES score gaps, but at the cost of lower predictive power for abilities that matter in college. Exams that do not measure ability precisely can reduce earnings if ability and college selectivity are complementary in human capital accumulation (Arcidiacono, 2004; Arcidiacono et al., 2016; Dillon and Smith, 2020).

Relatedly, not all low SES students will benefit from an exam reform that reduces the *average* SES test score gap. Although low SES students will be more likely to be in the "shifted in" group than in the "shifted out" group, some low SES students may have been better off under the old exam. For example, if SES test score gaps fall because the exam becomes a noisier measure of ability, this will tend to reduce scores for high-ability students in all socioeconomic groups. In particular, such a reform would restrict college choice for high-achieving low-income students (Hoxby and Avery, 2013; Dynarski et al., 2021), who may have particularly high returns to attending selective colleges.

A final consideration is that exam reforms can also alter the composition of a college's student body, which may generate spillover effects on an individual student's return to attending that college. For example, spillover effects could arise from peer effects in learning (Sacerdote, 2001) or if professors adjust their teaching level to the new class composition (Duflo et al., 2011). In this case, students' earnings outcomes may be affected even if they would have attended the same college program under both the pre- and post-reform exams. In most papers in the literature, college peer composition is either implicitly or explicitly held fixed in the authors' identification strategy.

Given the many dimensions of heterogeneity and the potential for spillovers, it is important to have empirical evidence on how exam reforms affect earnings outcomes. With this motivation, I turn to my analysis of the impacts of the Colombian admission exam reform.

2. Admission exam reform in Colombia

2.1. **Institutional background and data.** My analysis uses three administrative datasets from the Colombian higher education system. First, I use records from a national standardized exam called the ICFES, which all Colombian students are required to take to apply to college. The ICFES is analogous to the U.S. SAT exam, but it is taken by nearly all high school graduates. My main analysis uses individual administrative records from the testing agency that cover all 1998–2001 exam takers. These provide each student's test scores and high school affiliation, as well as several measures of SES. In supplementary analyses, I use data on 2002–2007 ICFES exam takers to examine longer-run effects, but the testing agency did not collect SES measures for most of these cohorts.

Second, the Ministry of Education provided enrollment and graduation records for the near universe of colleges in the country. Colombia's college system consists of public and private institutions with varying selectivity and degree offerings. The Ministry's records cover students who enrolled in nearly all of these colleges in 1998–2012.[4] The data include each student's institution, program of study, dates of entry and exit, and graduation outcome.

Finally, I use tax data from the Ministry of Social Protection, which contain monthly employment and earnings for any college enrollee working in the formal sector. My labor market data cover two time periods: 2008–2012 and 2017. The 2008–2012 data allow me to measure earnings 10–11 years later for each of the ICFES exam cohorts in my main sample (1998–2001), and these data include the number of employment days in each month. With the 2017 data, I observe earnings at different lengths of time since the ICFES exam (16–19 years later), and the data from this year do not include the number of employment days.

I link the admission exam, college, and earnings records using names, birthdates, and ID numbers. Appendix C.1 provides details on the coverage of each dataset and the merge.

2.2. **Variable definitions.** I define three measures of income and socioeconomic inequality from information collected at the time students took the admission exam:[5]

(1) Gaps between the top and bottom quartiles of the family income distribution;
(2) Gaps between students with college and primary (or less) educated mothers;
(3) Gaps between students who attended high and low ranked high schools.

My main outcome variables are ICFES exam scores, college selectivity, college graduation, and earnings. I convert raw ICFES scores to percentiles within a student's exam cohort. I define college selectivity as a school's mean admission score across all pre-reform students and all subjects, and convert it to percentiles in the same manner. This is a fixed measure of selectivity that is common in the literature (Dale and Krueger, 2002; Hoxby, 2009), and it provides a natural link between admission exam design and college outcomes. My graduation variable is an indicator for graduating from any college in the Ministry of Education records by 2012. I use two earnings variables: 1) log average *daily* earnings measured 10–11 years after the ICFES exam;[6] and 2) log *annual* earnings in 2017.

Table 1 shows summary statistics for these variables by SES group. I observe 1.6 million high school graduates who took the ICFES exam in 1998–2001. Most of my analysis focuses on the 612,949 students who enrolled in college. Columns (C)–(G) show large gaps in all outcomes between SES groups. For example, students from the top income quartile scored 26 percentile points higher than bottom quartile students on the pre-reform math

[4] Individual-level data from the Ministry of Education are not available before 1998.
[5] See Appendix C.1 for details on the definition of these SES measures.
[6] I compute average daily earnings by dividing total annual earnings by the number of formal employment days in the year, demeaning by exam cohort and year, and averaging across the two years.

component of the ICFES exam (column C). In the pre-reform cohorts, top quartile students attended colleges ranked 20 percentile points higher on average (column D), and they were 14 percentage points more likely to graduate from college conditional on enrolling (column E). Lastly, the gap in daily earnings between top and bottom quartile students was 44 percent measured 10–11 years later (column F), and this gap widened to 74 percent measured by annual earnings in 2017 (column G).

2.3. **The 2000 ICFES exam reform.** The ICFES exam was first administered in 1968 with the aim of supporting college admissions. As the exam gained widespread coverage in the 1980s, the government began using its results to evaluate high schools. In the mid 1990s, policymakers concluded that the exam was poorly designed for the dual objectives of college admissions and high school accountability (Buitrago and Blanco, 2002). Critics argued that the exam rewarded test prep and memorization more than learned material. Thus policymakers believed the exam was not well linked to high school curricula, and that scores were biased toward high SES students who had greater access to test prep services.

To address these concerns, the testing agency implemented a complete redesign of the ICFES exam. The style of questions changed with the goal of testing "competencies" rather than "content." The new exam often asked questions based on graphs or reading passages that were provided, while the old exam frequently required students to recall facts or formulas that they learned elsewhere. The testing agency changed the subjects that appeared on the exam to try to create better alignment with high school and college curricula. For instance, the overhaul converted two math components ("aptitude" and "knowledge") into a single math test, and it added a foreign language test. Further, the new ICFES had four options for each multiple choice question as opposed to five, and it had fewer questions overall (*El Tiempo*, 1999). Appendix C.2 describes the ICFES overhaul in detail and provides examples of questions from the pre- and post-reform tests.

The new ICFES exam debuted in 2000 after more than five years of psychometric research and testing. In preceding months, the overhaul was widely publicized in media outlets including *El Tiempo*, the leading Colombia newspaper. In its objective and publicity, the redesign of the exam was similar to the overhaul of the U.S. SAT exam in 2016.

2.4. **Reduction in SES test score gaps.** Figure 1 shows that the ICFES reform significantly reduced test score gaps between high- and low-income students. The height of each bar is the difference in the mean score percentile for students in the top and bottom income quartiles. Throughout the paper, I focus on the six subjects that appeared on both the old and new exams: math, language arts, biology, chemistry, physics, and social science (see Appendix Table A2). In the pre-reform cohorts (1998–1999), the test score gaps were roughly 27 percentile points in each subject. The reform had a dramatic impact on math and physics scores, with test score gaps falling by 50 percent in the post-reform cohorts (2000–2001). There were smaller reductions in language, biology, and chemistry test score gaps, and only minor effects in social science.

Table 2 shows that the reform also reduced test score gaps defined by other measures of SES. Column (A) displays mean test score gaps across the six subjects in pre-reform cohorts. These gaps ranged from 24 to 31 percentile points using family income, mother's education, and high school rank as measures of SES. Columns (B)–(G) show the change in test score gaps between the pre- and post-reform cohorts by exam subject. The reform reduced gaps in math and physics scores by roughly 50 percent for each SES measure. The decline in test score gaps was more modest in other subjects, and the social science gap declined by only 1–3 percentile points.[7] Thus the testing agency achieved its goal of reducing the relationship between socioeconomic status and exam scores.

2.5. **Reduction in exam validity.** A second objective of the ICFES reform was to design a better measure of abilities that predict college success. I examine how the reform affected the predictive power of the exam in two ways.

First, I follow testing agencies' standard way of measuring exam validity, which correlates test scores and college outcomes. I use three college outcomes:

(1) Scores on a field-specific college *exit* exam that students take prior to graduation.[8]
(2) An indicator for graduating from any college in the Ministry of Education data.
(3) First-year GPA; I do not observe GPA in my administrative data, but I use a subsample of students in the 2000–2004 enrollment cohorts at a flagship university for which I have transcript records. See Appendix C.1 for details.

To reduce the influence of a student's college choice, I follow the standard practice of using residuals from regressing both test scores and outcomes on college dummies (e.g., Kobrin et al., 2008). Thus the correlation coefficients reflect only within-college variation.

Second, I use social science scores as a relatively fixed measure of ability. To measure exam validity, one would ideally observe students' performance on a similar test taken earlier in high school. Such data are typically not available to testing agencies, and I also do not have access to an earlier measure of ability. As an alternative, I show how the reform affected the correlation of social science scores with other subject scores. Although this analysis is imperfect, the reform had a measurably smaller effect in social science (Table 2), and these scores partially measure a common component of ability.

[7] The reform also reduced the gender gap in ICFES scores in several subjects, but I do not find that it significantly affected gender gaps in college selectivity or other outcomes.

[8] The Colombian exit exam is called *Saber Pro* (formerly ECAES), and it is also administered by the ICFES testing agency. My data include all students who took the exit exam in 2004–2011. The exit exam is now a national requirement for graduation, but it was optional during this time period (MacLeod et al., 2017).

Panel A of Table 3 shows that the ICFES reform reduced the exam's predictive power for both college outcomes and social science scores. Column (A) shows how the (within-college) correlation of exit and admission exam scores changed from the pre- to the post-reform cohorts.[9] This correlation fell by about 0.15 points in math and physics—the subjects with the largest reductions in SES test score gaps. This is a 40 percent decrease in validity from the mean pre-reform correlation of 0.39. The reform induced modest reductions in exit exam validity in language arts and biology, and had little effect in chemistry and social science.

Columns (B)–(D) of Panel A show that the reform also reduced the exam's validity for college graduation, first-year GPA, and social science performance. The correlation between math/physics scores and graduation fell by 0.06 points, or roughly 50 percent of the pre-reform mean. Reform effects on graduation validity are mixed for other subjects. The GPA effects are underpowered because few students in my transcript data took the pre-reform exam. Nonetheless, the math and physics exams became weaker predictors of first-year GPA. Lastly, the reform drastically reduced the correlation of social science scores with scores in other subjects, particularly math and physics.

Panel B of Table 3 shows that the reform also reduced the exam's predictive power *within* SES groups. This panel is identical to Panel A, except I residualize all variables on covariates for family income, mother's education, and high school rank. I find similar results on within-SES validity; most coefficients are negative and only slightly smaller in magnitude than those in Table 3. This shows that the decline in the exam's predictive power did not arise solely from the reduction in its correlation with observable measures of SES.[10]

These findings suggest that SES test score gaps fell primarily *because* of the decrease in the exam's validity. Although the reform aimed to make the exam a better predictor of college success, structural changes that reduced the exam's information content outweighed any improvements in the style of questions. The new exam had fewer questions overall, and each question was more "guessable" because it had fewer multiple-choice options. Both of these changes cause a larger proportion of the variance in scores to come from randomness (Riehl and Welch, 2023).[11] This effect was particularly pronounced in math because the reform converted two separate test components into one.

The rest of the paper asks whether the reduction in test score gaps also reduced inequality in college selectivity and labor market outcomes.

[9] I normalize variables in Table 3 so that regression coefficients can be interpreted as correlation coefficients.

[10] Appendix Figures A2–A3 show that the ICFES reform had the largest effects on scores at the top and bottom of the ability distribution. In other words, students with the highest ability experienced the largest score decreases, and students with the lowest ability experienced the largest score increases.

[11] Anecdotally, many test takers found the new ICFES exam to be quite difficult, which implies that guessing was likely to be an important contributor to the total variance in scores (*El Tiempo*, 2000).

3. Identification and empirical strategy

3.1. **College admission markets.** To identify the effects of the ICFES exam reform on longer-run outcomes, I exploit three features of the Colombian higher education system. First, Colombian students typically attend college close to home. About 75 percent of tertiary students attend a school in the city of their birth (Saavedra and Saavedra, 2011), and students who are from rural areas often attend college in the capital city of their region.

Second, the local public university is the most desirable college for many students. During the period of my data, Colombia had roughly 50 public universities including a flagship school in the capital city of most regions. Public schools are much less expensive than comparable private colleges and give tuition discounts to low SES students. Flagships are typically the region's largest and most prestigious college; in some regions, more than one-third of all college students attend the flagship school. As a result, admission to public universities is highly competitive. Colombia also has a large network of private colleges including several small elite universities, but admission to these schools is much less competitive because they are either less prestigious or too expensive for most students to attend.[12]

Third, public universities varied in whether or not they used the ICFES exam in admissions. As in the United States, college admissions in Colombia are decentralized; students apply separately to each college, and schools determine their admission criteria. Most public universities consider *only* ICFES scores in admissions. These schools compute weighted averages of ICFES subject scores and admit the highest ranking students up to a quota. Admissions are at the school/major level, and the weights on each subject score usually vary across programs. But some public universities do not use ICFES scores. Most commonly, these colleges require applicants to take the school's own exam and use these scores as the sole admission criterion.[13] For example, the flagship school in Cali (Universidad del Valle) used the ICFES exam, while the flagship in Bogotá (Universidad Nacional de Colombia) administered its own entrance exam.

These factors created geographic variation in the stakes of the ICFES exam reform. The reform was more consequential in areas where the local public university used only ICFES scores for admissions. Since public universities are large and highly selective, changes in their admission outcomes would also affect enrollment at other nearby colleges. The reform mattered less in areas where the local public university used other admission criteria.

[12] Financial aid markets were essentially non-existent during the period of my data, and tuition was roughly five times higher at top private schools (Riehl et al., 2019). The top private college (Universidad de Los Andes) admitted roughly half of its applicants during this time, while the admission rate at the flagship university in Bogotá (Universidad Nacional de Colombia) was near ten percent.

[13] Some colleges also consider other factors like high school grades or interviews. One explanation for the variation in admission methods is that it is costly to administer a separate exam. Colleges that rely on the ICFES exam tend to be in less-populated regions, and thus have lower budgets.

3.2. **Definition of treated areas.** To capture this variation, I define two geographic measures of reform "treatment" using the pre-reform admission criteria at public universities.

First, I define *region-level* treatment based on public university admission methods in the administrative region (*departamento*) where a student attended high school. I define a student as "treated" if they went to high school in a region where public universities used only ICFES exam scores in admissions. "Control" students are from regions where public universities use their own entrance exams or other admission criteria. Colombia has 33 administrative regions, and in most cases this classification is unambiguous because the region contained only one public university and/or admission method. In the three regions with mixed admission criteria, I define treatment as the (enrollment-weighted) modal admission method.[14] For the sparsely-populated regions with no universities, I define treatment using the closest public university to the region's capital city.[15] Appendix Table A3 provides details on public university admission methods and this region-level treatment variable.

Second, I define *municipality-level* treatment based on the pre-reform propensity to enroll in ICFES exam universities in each of Colombian's roughly 1,000 municipalities (*municipios*). This treatment variable equals the ratio of: 1) the number of 1998–1999 exam takers in a municipality who enrolled in a public university with ICFES-based admissions; to 2) the number of 1998–1999 exam takers in that municipality who enrolled in any college. This is a continuous measure of treatment that reflects exposure to universities that used the ICFES exam in the municipality where a student attended high school.

Figure 2 shows the region- and municipality-level treatment variables. Black dots in both panels represent public universities with ICFES admissions, and white dots show public universities with other admission criteria. Panel A shows the binary region-level variable, with treated regions shaded dark red and control regions shaded light yellow.[16] Panel B shows municipality-level treatment, with darker colors reflecting higher pre-reform enrollment rates in ICFES exam universities. Enrollment rates range from near zero in some municipalities to over 50 percent in urban areas near ICFES exam universities. The correlation between the two treatment measures is 0.76. The mean of the municipality-level treatment variable is 36 percent in treated regions and five percent in control regions.

[14] Specifically, Bogotá is a control region because only 17 percent of public university students attended schools with ICFES exam admissions. Caldas and Valle del Cauca are treated regions because 74 percent and 92 percent of public university students attended schools that used the ICFES exam.

[15] My results are similar when I exclude small regions because they have few college students. Results are also similar when I exclude regions with mixed admission criteria (Bogotá, Caldas, and Valle del Cauca), although they are less precisely estimated. See Appendix Tables A14 and A15.

[16] The capital city of Bogotá is its own administrative region. Figure 2 does not display the Caribbean island region of San Andrés y Providencia. I define both of these areas as control regions.

3.3. **Difference-in-differences specification.** My empirical approach combines geographic variation in the stakes of the ICFES exam with the timing of exam reform. My benchmark specification is a standard difference-in-differences regression,

$$y_{igt} = \gamma_g + \gamma_t + \theta(\text{Treated}_g \times \text{Post}_t) + u_{igt}, \tag{1}$$

where y_{igt} is an outcome for student i who attended high school in geographic area g and took the ICFES exam in cohort t. The regression includes geographic area dummies, γ_g, exam cohort dummies, γ_t, and the interaction between the treatment variable, Treatment_g, and a dummy for post-reform exam cohorts, Post_t. I run regressions with both region- and municipality-level treatment variables, and use region- and municipality-level dummies correspondingly. I cluster standard errors at the region level in all regressions.[17]

The coefficient of interest, θ, measures how outcomes changed with the reform in treated areas relative to control areas. In regressions with region-level treatment, θ reflects the average change in outcomes in treated regions relative to control regions. In municipality regressions, I normalize Treated_g so that θ measures the effect of a one standard deviation increase in pre-reform enrollment at ICFES exam universities.

3.4. **Identification assumptions and balance tests.** The main identification assumption is the usual difference-in-differences requirement of parallel trends. In this context, parallel trends means that college and labor market outcomes would have evolved similarly in treated and control areas in absence of exam reform. This section shows evidence of parallel trends in observable student traits, high school graduation rates, and macroeconomic conditions.

Panel A of Table 4 shows student characteristics in treated and control areas. Columns (A) and (B) display means for pre-reform exam takers in treated and control regions. My region-level treatment classifies 20 regions as treated and 13 as control. Control regions are more densely populated on average, with roughly 50 percent more exam takers per cohort. Control region students come from more socioeconomically advantaged backgrounds as measured by family income, parents' education, and attendance at top high schools.

While treated and control students have different characteristics, I find little evidence of differential trends in student traits across cohorts. Columns (C)–(F) display θ coefficients and standard errors from separate estimations of equation (1) using the dependent variable in the first column. Columns (C)–(D) show estimates with the region-level treatment variable, and columns (E)–(F) use municipality-level treatment. Point estimates are small and are statistically significant only for mother's education. I find precise zeroes in balance tests that combine all individual characteristics into a single index using the predicted values from a

[17] Appendix Tables A12 and A13 show that the main results in this paper are robust to the wild t bootstrap procedure recommended by Cameron et al. (2008) for specifications with a relatively small number of clusters.

log earnings regression. These results suggest that changes in cohort composition were not systematically related to exposure to ICFES exam universities.

Panel B presents evidence of parallel trends in economic conditions. Colombia experienced a severe recession in the late 1990s, and any heterogeneous effects of this recession could have caused divergent trends in college choices. To explore this possibility, Panel B present estimates of labor force participation and unemployment rates using region-level data from a Colombian household survey (*Encuesta Nacional de Hogares*).[18] Treated and control regions had relatively similar pre-reform employment conditions, and there is little evidence of differential trends during the recession.

Figure 3 corroborates this finding by plotting trends in high school graduation and unemployment rates from labor market surveys. Both panels display mean outcomes in treated and control regions, with a vertical line separating pre- and post-reform periods. Panel A shows high school graduation rates for cohorts defined by the year individuals turned 17—the most common age at which students take the college admission exam. High school graduation rates climbed rapidly in the 1990s and are higher in control regions, but they followed similar paths in the two areas. Panel B shows that unemployment rates rose from ten percent in the mid 1990s to over 16 percent by the year 2000, yet treated and control areas experienced similar trends through the downturn. A caveat is that I cannot rule out moderately-sized differences in employment trends due to volatility.

3.5. **Behavioral responses to exam reform.** Panel C of Table 4 presents evidence on three potential behavioral responses to the ICFES reform: exam retaking, the likelihood of attending any college, and geographic mobility. Such responses are potential causal effects of the reform, but they would affect the interpretation of my results.

The reform had no significant effect on exam retaking or the probability of enrolling in any college. Although students in treated regions were more likely to retake the admission exam (columns A–B), changes in retaking rates are unrelated to the region- and municipality-level treatment variables (columns C–F). Overall college enrollment rates were similar in treated and control regions and did not diverge significantly with the reform. Appendix Table A4 also shows little evidence of differential changes in these outcomes by SES group. A potential explanation for these null results is that Colombia has many non-selective colleges where students can enroll if they are not admitted to top colleges. This is consistent with work that finds that admission policies at selective universities have little effect on the extensive margin of college enrollment (Hinrichs, 2012; Daugherty et al., 2014).

There is some evidence that the reform affected student mobility across regions. There are no significant changes in the mean distance between a student's high school and college, but I

[18] Individual and municipality-level household survey data are not available prior to 2001.

find a 1–2 percentage point increase in the probability that treated students stayed in region for college. Similarly, I find a marginally significant effect on the fraction of college enrollees who attended a public university. Thus there may have been an increase in student body size at ICFES exam universities relative to other public universities, but the magnitude of this effect is small relative to normal fluctuations in cohort size.[19] Further, many Colombian students do not enter college right after high school (de Roux and Riehl, 2022), so any class size effects were spread out over multiple cohorts. Thus I believe that overcrowding or other cohort size effects are unlikely to explain my main results.

4. Exam reform effects on college selectivity, graduation, and earnings

4.1. **Effects on the SES college selectivity gap.** I first examine the effects of the ICFES exam reform on access to selective colleges. For this I modify the difference-in-differences regression (1) to estimate changes in the *gap* in college selectivity between high and low SES students. I use a measure of college selectivity, Q_c, as the dependent variable, and I interact all right-hand side variables with a dummy for high SES students, X_i. This yields the triple differences regression:[20]

$$Q_c = \gamma_{gt} + \Big(\gamma_g + \gamma_t + \theta^q(\text{Treated}_g \times \text{Post}_t)\Big)X_i + u_{icgt}. \tag{2}$$

The coefficient of interest, θ^q, measures the change in the SES college selectivity gap in treated areas relative to control areas. As above, Q_c is a fixed measure that equals a college's percentile rank based on mean pre-reform scores, so θ^q is in percentile units. The reform made low SES students more competitive for admission to selective colleges on average. Thus the prediction is that the reform should also reduce the SES college selectivity gap in areas where selective colleges use the ICFES exam in admissions ($\theta^q < 0$).

Table 5 presents estimates of θ^q from equation (2). Panel A uses the region-level treatment variable with region dummies (γ_g and γ_{gt}). Panel B uses municipality-level treatment with municipality dummies. Each row presents a coefficient from a separate regression that defines high SES, X_i, by family income, mother's education, or high school rank. Column (A) displays my benchmark estimates of θ^q coefficients from equation (2).

The results in column (A) show that the reform reduced gaps in college selectivity between high and low SES students. The first row of Panel A shows that the college selectivity gap by family income decreased by 2.4 percentile points in treated regions relative to control regions. This is roughly a ten percent decline from the pre-reform mean college selectivity gap

[19] The largest estimate in Table 4 equates to 90 extra students per cohort at each public university. The mean cohort size at these schools was about 2,000 students, with a within-school standard deviation of 400.
[20] Equation (2) is equivalent a two-step estimation procedure: 1) estimate the college selectivity gap in area g and cohort t; and 2) use this gap as the dependent variable in the difference-in-differences regression (1). See Card and Krueger (1992) for an example of this two-step specification.

of 20 percentile points (Table 1). The magnitude is similar using municipality-level treatment (Panel B); a one standard deviation increase in pre-reform exposure to ICFES exam universities is associated with a 1.5 percentile point decrease in the college selectivity gap by family income. The effects are smaller in magnitude using mother's education and high school rank as measures of inequality, though in all cases college selectivity gaps declined.

Figure 4 shows that the college selectivity gap in treated areas declined sharply in the first cohort after the reform. Panel A plots the mean college selectivity gap between top and bottom income quartile students in treated regions (dashed line) and control regions (solid line). The left and right vertical axes are shifted so that the treated and control region gaps match in the last cohort before the reform. There is little evidence of differential pre-trends in the two pre-reform cohorts (1998–1999), and the college selectivity gap in treated regions fell by about two percentile points in the 2000 exam cohort. Panel B shows similar results using an event-study version of the municipality-level regression based on equation (2), which estimates separate θ_t^q coefficients for each exam cohort t (omitting the 1999 coefficient).[21]

4.2. **Robustness tests for college selectivity effects.** A potential concern is that treated areas are more populated and closer to universities than control areas (Table 4), and thus there may have been divergent trends in college selectivity. Urban areas may have been more affected by the contemporaneous recession, which may have differentially affected students' colleges choices. Colombia's rising educational attainment may have led to different enrollment trends in areas that were closer and farther from public universities.

To test these potential concerns, I add controls to the benchmark regression to restrict identification to municipalities with similar populations and proximities to public universities. Specifically, I divide municipalities into ten groups based on population, with a separate eleventh group for the three largest cities (Bogotá, Medellín, and Cali). I also define ten municipality groups based on distance to the nearest public university.[22] I then estimate equation (2) including dummies for these municipality groups fully interacted with cohort and SES dummies.[23] With these controls, college selectivity effects are identified within groups of 14 municipalities on average, where all municipalities in a given group have similar populations and proximities to public universities.

Column (B)–(C) in Table 5 show that the college selectivity effects persist with the addition of these controls. The decline in the family income gap is *larger* in magnitude when I restrict identification to municipalities with similar populations (column B). The effects

[21] Appendix Figure A4 presents graphs similar to Figure 4 for college selectivity gaps by mother's education and high school rank. Results are similar, but there is evidence of a pre-trend in the mother's education gap.
[22] Appendix Table A5 shows the population and proximity groups. I define groups using Ward's method for hierarchical clustering, which minimizes the total within-cluster variance given the choice of K clusters.
[23] For example, the regressions in column (C) of Table 5 include $\gamma_{K(g)J(g)xt}$ dummies, where $K(g)$ defines population groups, $J(g)$ defines proximity groups, x defines SES groups, and t defines exam cohorts.

on college selectivity gaps by mother's education and high school rank remain negative, although the magnitudes fall and standard errors increase. I find similar results when I include both population and proximity controls (column C). This suggests that the declines in college selectivity gaps are not attributable to differential trends between urban and rural municipalities, or between areas that differ in distance to public universities.

4.3. **Effects on graduation and earnings.** Table 6 presents my main results on the effects of the ICFES reform on college graduation rates and earnings. The table displays estimates of θ from the standard difference-in-differences regression (1) using four dependent variables: an indicator for college persistence, defined as still being enrolled one year after starting college (column A); an indicator for college graduation by 2012 (column B); log daily earnings measured 10–11 years after the ICFES exam (column C); and log annual earnings in 2017 (column D). Panel A presents estimates using the region-level treatment variable, and Panel B uses the municipality-level treatment variable. In each panel, the first row shows the average effects measured across all college enrollees. The remaining rows show coefficients from regressions for separate groups of high and low SES students.

The reform lowered average graduation rates and post-college earnings in treated areas relative to control areas. The first row of Panel A shows that college graduation rates declined by 1.5 percentage points in treated regions relative to control regions (column B). The effect on college persistence is similar in magnitude (column A), suggesting that the decline in graduation rates came primarily from students dropping out in their first year. The reform also led to 1.3 percent decrease in average daily earnings measured one decade later (column C). This negative earnings effect was not transitory, as the estimate on log annual earnings in 2017 is even larger in magnitude (column D). Panel B shows that the effects of the ICFES reform are similar using municipality-level treatment; these coefficients are smaller than those in Panel A—consistent with the smaller magnitude of the college selectivity effects in Table 5—and they are significant at $p < 0.05$ for all outcomes.

Figure 5 shows that average graduation rates and earnings decreased in the first cohort after the ICFES reform. This figure is similar to Figure 4; Panels A and C plot mean graduation rates and earnings in treated and control regions, and Panels B and D plot municipality-level event study estimates based on equation (1). There is no evidence of differential pre-trends in either outcome, and in all cases outcomes decline immediately in the 2000 exam cohort. These sharp declines suggests that these effects were driven by the ICFES reform rather than by longer-term macroeconomic trends.

The other rows of Table 6 show that *both* high- and low-SES students experienced declines in graduation rates and earnings. For example, the second and third rows of Panel A show that the ICFES reform reduced graduation rates for students in the top income quartile by

2.7 percentage points, and it reduced graduation rates for bottom quartile students by one percentage point (column B). Similarly, the reform reduced daily earnings measured 10–11 years later by 1.8 percent for top quartile students, and by 1.4 percent for bottom quartile students (column C). These patterns are similar using the municipality-level treatment variable (Panel B) and other measures of SES (lower rows). The SES-specific estimates are not always statistically significant as standard errors are larger, but the coefficients are negative for all SES groups, outcomes, and treatment variables.

4.4. **Robustness tests for graduation and earnings effects.** Appendix Tables A6–A9 show that the graduation and earnings results in Table 6 are robust to including controls for population and college proximity. These regressions are similar to those in columns (B)–(C) of Table 5 in that I include dummies for municipality population and proximity groups interacted with exam cohort dummies. These controls have little effect on the point estimates, although the coefficients become statistically insignificant in some cases.

An important caveat is that I have a short pre-period to examine parallel trends in Figure 5 because my administrative datasets are only available beginning in 1998. As an alternative, I test for parallel trends in college graduation rates and earnings using a Colombian household survey called the *Gran Encuesta Integrada de Hogares* (GEIH). For this analysis, I measure outcomes in 2015–2019 GEIH survey data, and I compare *birth cohorts* that were likely to have taken the ICFES exam before or after the 2000 reform. Appendix Figure A5 displays birth cohort event studies for graduation rates and earnings using GEIH data, and Appendix Table A11 presents the corresponding regression results.

I do not find evidence of diverging pre-trends in GEIH data, although this analysis is underpowered and thus only suggestive. The event study coefficients in Appendix Figure A5 are noisy, but I do not find systematic pre-trends in college graduation rates or log hourly earnings in the birth cohorts that would have mostly taken the pre-2000 ICFES exam (1970–1980).[24] Appendix Table A11 shows some evidence that, in subsequent birth cohorts, college enrollees' mean graduation rates and earnings declined in treated regions relative to control regions. The GEIH graduation and earnings coefficients are similar in magnitude to those in columns (B) and (D) of Table 6, although the earnings estimates are mostly insignificant. Notably, I find no evidence of negative earnings effects for high school graduates who did *not* attend college, which suggests that my results are not driven by correlated economic shocks.

The ICFES reform may have also caused a small decrease in the rate of formal employment. I only observe earnings for individuals who work at firms that are registered with the Ministry of Social Protection.[25] Appendix Table A4 shows estimates from equation (1) in which the

[24] Roughly 90 percent of ICFES exam takers in my administrative data are between the ages of 16 and 19.

[25] These formal workers comprise 56 percent of my sample for the 2008–2012 earnings data. Individuals who do not appear in the earnings data can be informally employed, unemployed, or out of the labor force.

dependent variable is an indicator for appearing in the earnings data 10–11 years after the ICFES exam. Estimates are insignificant using region-level treatment, but a one standard deviation increase in municipality-level treatment is associated with a one percentage point decrease in formal employment. This is another potential effect of the reform, but it is unlikely to change the sign of the earnings estimates in Table 6 because wages are typically much lower in the informal sector. Consistent with this, Appendix Table A10 shows that I continue to find negative effects when I use daily earnings in levels (including zeroes) or the inverse hyperbolic sine of earnings as dependent variables.

4.5. **Effects in later cohorts.** An important question is whether the negative effects of the ICFES reform on graduation rates and earnings were transitory or persistent across exam cohorts. For example, the reform may have only had short-run effects on SES gaps in test scores if high-SES students learned how to better prepare for the new exam over time. Even if the test score impacts were long-lasting, the effects on student outcomes may have been temporary if colleges were able to adjust their teaching practices or support services.

Appendix Table A16 shows the longer-run effects of the ICFES reform on test scores, college selectivity, graduation rates, and earnings. This table uses similar specifications as in Tables 2, 5, and 6, but the sample includes the 1998–2007 exam cohorts. Since the ICFES testing agency did not collect data on family income and parental education in all of these cohorts, these analyses use students' high schools to define a measure of SES.

In short, the reform effects on test scores and student outcomes persisted up through the 2007 cohort. SES gaps in test scores remained significantly lower on the post-reform exams up through the 2007 cohort (see also Appendix Figure A6). This finding corroborates the evidence from Section 2 that the reduction in SES test score gaps was driven primarily by noise, which may have made it harder for students to influence their scores on the new exam through test prep. Consistent with the test score effects, the reform led to a persistent decline in the SES college selectivity gap in treated regions relative to control regions. Lastly, mean graduation rates and earnings in treated regions also remained lower than in control regions up through the 2007 cohort.[26]

4.6. **Magnitudes.** The effects of the ICFES reform on socioeconomic diversity at selective colleges were similar in magnitude to the impacts of U.S. affirmative action policies. A back-of-the-envelope calculation shows that the estimates in Table 5 are equivalent to one in 20 low SES students moving to the next most selective college in their region, and one

[26] A caveat is that I only observe graduation outcomes up through 2012, and I measure earnings in 2017 for every cohort. The reform effects in later cohorts are not statistically significant given higher standard errors, but the point estimates do not change much relative to those for the 2000–2001 cohorts.

in 20 high SES students moving down a selectivity tier.[27] The magnitude of this effect is roughly similar to the impact that affirmative action had on racial diversity at University of California schools in the 1990s (Bleemer, 2022), but it is much smaller than the scale of affirmative action policies in Brazil (Machado et al., 2022).

The reform's impacts on graduation rates are plausible given the significant reallocation of students across colleges. Table 6 shows that the reform reduced college graduation rates by 1.5 percentage points measured across all students. As a benchmark, suppose that the graduation effects were driven *only* by the students who switched colleges as a result of the reform. In the scenario from the previous paragraph, this would imply that the reform caused three out of 10 college switchers to drop out.[28] This is a large effect, but it is within the range of estimates from research on college selectivity. For example, Goodman et al. (2017) find that enrollment in the Georgia State University System increases the likelihood of bachelor's degree attainment by 41 percentage points. The reform may have also had spillover effects on the graduation rates of students whose college choice was unaffected.

The graduation and earnings estimates in Table 6 are roughly similar in magnitude, suggesting that reductions in educational attainment are an important mechanism for the earnings effects. If the decrease in earnings was attributable *only* to students who were induced to drop out, the estimates would imply a college earnings premium of roughly one log point. This is larger than the cross-sectional college earnings premium in my data (0.62 log points), but many papers find that causal estimates of the returns to education are larger than OLS estimates (Card, 2001). Of course, the earnings effects could also be due in part to changes in college selectivity that did not affect degree attainment.

The negative outcomes for high SES students are consistent with reductions in their college selectivity. There is a large literature documenting that increases in college selectivity often lead to higher graduation rates and earnings (Arcidiacono and Lovenheim, 2016). In Colombia, as in many countries, selective universities spend more per student than other colleges (Appendix Table A1 and Figure A1), and research finds that financial resources matter for graduation rates (Bound et al., 2010; Deming and Walters, 2017). Thus the results for high SES students suggest that decreases in their college selectivity reduced their likelihood of earning a degree, which harmed them in the labor market.

[27] The ICFES reform reduced the SES college selectivity gap by about two percentile points (Table 5). There are five colleges per treated region on average, so moving to the next most selective college corresponds to a 20 percentile point increase in selectivity. Thus a two percentile point decline could arise if one in 20 low SES students moved to a more selective college, and one in 20 high SES students moved to a less selective college. The true effects of the ICFES reform were more complex since it affected admissions at many programs.

[28] For example, if the reform caused five percent of high SES students to attend a less selective college and thirty percent of these switchers were induced to drop out, then the graduation rate for high SES students would decline by 1.5 percentage points (0.05*0.3).

The reductions in graduation rates and earnings for low SES students are more striking because these individuals experienced *increases* in college selectivity on average. I provide evidence on the mechanisms for this result in the next section.

5. Mechanisms

5.1. Heterogeneity by field of study. To understand the mechanisms for my graduation and earnings results, I begin by examining heterogeneity by field of study. The effects of the ICFES reform were largest on the exam's math and physics components, for which both SES test score gaps and the predictive power of scores declined by about 50 percent (Tables 2 and 3). The effects on test score gaps and validity were more modest in other subjects. Thus one might expect that the reform's impacts on student outcomes were most pronounced in college programs where math and physics skills were important for earning a degree.

Table 7 examines heterogeneity using three different categorizations of fields. In Panel A, I define three field groups based on the college programs that students enrolled in: STEM, business, and health/humanities.[29] Panel B uses the same three groups, but fields are defined by the college program that students listed as their top choice at the time they took the ICFES exam. Finally, Panel C groups college programs based on the importance of math and physics skill for graduation. For this, I regress an indicator for degree completion on pre-reform ICFES subject scores for *each* college program, and then divide programs into quartiles based on the sum of the math and physics coefficients from this regression. Programs in the top quartile are those in which math and physics scores had the highest predictive power for graduation in the pre-reform cohorts.

Columns (B)–(C) of Table 7 show that math and physics skill is most important in STEM and business fields. These columns display math and physics coefficients from a regression of degree completion on the six pre-reform ICFES subject scores for each field, with scores normalized to standard deviation units.[30] In STEM programs, the correlation between pre-reform math scores and graduation rates (conditional on other subject scores) is 0.074, and the physics score correlation is 0.062. Math scores are also strongly related to graduation in business, but physics scores have less predictive power. These conditional correlations are much lower in health and humanities programs (0.032 and 0.030). I find similar patterns using students' desired fields at the time of the exam (Panel B), although the differences are

[29] This categorization uses the Ministry of Education's classification of programs into field of study areas. STEM includes natural sciences and engineering programs. Business includes economics. Health and humanities includes programs all other fields, including education, law, and the qualitative social sciences. This categorization creates three groups of roughly equal size based on pre-reform enrollment (see column A).

[30] This is the same regression that defines the quartiles in Panel C, except I pool across all programs in a field. I normalize variables to standard deviation units, so coefficients can be interpreted as the pre-reform correlations of math and physics scores with graduation rates controlling for other subject scores.

more modest. Math and physics scores vary significantly in their predictive power across the program groups in Panel C, which is a mechanical effect of how I define quartiles.

Column (D) of Table 7 shows that the ICFES reform did not have a significant effect on the proportion of students in each field. This column displays θ coefficients from the difference-in-differences regression (1). The regression sample includes all 1998–2001 college enrollees (regardless of field), and the dependent variables are indicators for choosing each field. The estimates in column (D) are all close to zero and statistically insignificant, which shows that the ICFES reform did not have differential effects on enrollment in each field in treated and controls regions. This suggests that the ICFES reform primarily affected the colleges that students attended rather than their fields of study.[31]

My main finding in Table 7 is that the negative graduation and earnings effects of the ICFES reform are concentrated in STEM and business fields. Columns (E)–(F) display θ coefficients from estimating equation (1) separately for each field, with college graduation and log daily earnings 10–11 years later as dependent variables. For students who enrolled in STEM programs, the reform reduced graduation rates by 3.1 percentage points, and earnings declined by nearly three percent (first row of Panel A). The reform reduced earnings by 1.6 percent in business, although this effect is not statistically significant. By contrast, the reform effects in health and humanities programs were close to zero. I find the same pattern using students' field preferences (Panel B), and the negative effects were similarly concentrated in programs where math and physics skills matter most for graduation (Panel C).

5.2. **Mismatch.** The results in Table 7 suggest that the ICFES reform led to mismatch between students and colleges on the basis of academic preparation, which can explain why earnings declined even for low SES students. Some research finds that graduation rates and earnings are higher for students whose own ability is close to that of their classmates, and that relative academic preparation is particularly important in STEM programs (Stinebrickner and Stinebrickner, 2014; Arcidiacono et al., 2016). My results are consistent with this finding in that the negative impacts of the reform were concentrated in quantitative fields like STEM and business. In these fields, math and physics scores were most informative for student success prior to the reform, and thus they were most affected by the loss of information from the exam redesign.

My results are not conclusive on whether the ICFES reform led to "mismatch" in the usual sense (Arcidiacono and Lovenheim, 2016), or whether student/college match quality declined for other reasons. Mismatch can arise because students have negative returns to college selectivity, or because students with particularly high returns to selective colleges

[31] I also do not find significant effects on field of study choice when I estimate these regressions separately for high- and low-SES students (see Appendix Table A17). The results in Panel B of Table 7 indicate balance with respect to students' field preferences at the time of the ICFES exam.

are not admitted. Further, the ICFES reform altered both the socioeconomic and academic diversity of colleges' student bodies, which may have impacted student outcomes through peer channels. My results do not distinguish between these different mechanisms since the exam reform had complex effects on student/college matches, even within the population of low SES students.

Regardless of the type of mismatch, my findings shows that information on student ability is important in college admissions. Although the ICFES exam redesign succeeded in reducing SES gaps in test scores and college selectivity, the decline in the exam's informativeness outweighed the benefits of increased college access for the average low SES student.

6. Conclusion

A large literature asks how attending a selective college affects an individual's career prospects (e.g., Dale and Krueger, 2002). This work helps to explain why families expend a great deal of energy on college admissions (Ramey and Ramey, 2010), as it explores the impacts of college choice from the perspective of individuals.

This paper explored college choice from a market perspective. It asked how the matching of students to colleges via an admission test affects the distribution of earnings. Using a natural experiment in Colombia, it showed that the earnings implications of an admission test depend on the exam's predictive power for student ability. Exams with low predictive power for these returns can lead to mismatch in the assignment of students to colleges. This was the case in Colombia, where an admission exam overhaul reduced test score gaps but also the information content of the exam. The result was that post-college earnings fell for both high and low SES students.

These results suggest that one should be cautious in extrapolating estimates from marginally admitted students to large-scale admission policies. Returns may differ for students whom a college deems to be qualified and for those who would not be admitted under status quo policies. Further, large-scale reforms can lead to spillover effects through a variety of peer channels, which can change the returns for any individual student.

Another takeaway is that increasing "fairness" in admission tests does not necessarily reduce inequality in labor market outcomes. Reforms that lower test score gaps may not help disadvantaged students if the exam becomes a worse measure of skills that matter in college. A growing number of colleges have adopted test-optional admission policies or done away with admission exams altogether. The results in this paper show that it is important for these schools to use other admission criteria that identify which students are academically prepared to succeed.

References

Abdulkadiroğlu, A., P. Pathak, and A. Roth (2005). The New York City high school match. *American Economic Review 95*(2), 364–367.

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.

Anelli, M. (2020). The returns to elite university education: A quasi-experimental analysis. *Journal of the European Economic Association 18*(6), 2824–2868.

Angrist, J., D. Autor, and A. Pallais (2022). Marginal effects of merit aid for low-income students. *The Quarterly Journal of Economics 137*(2), 1039–1090.

Angrist, J. D., P. A. Pathak, and R. A. Zárate (2019). Choice and consequence: Assessing mismatch at chicago exam schools. National Bureau of Economic Research.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1), 343–375.

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.

Arcidiacono, P. and J. L. Vigdor (2010). Does the river spill over? estimating the economic returns to attending a racially diverse college. *Economic Inquiry 48*(3), 537–557.

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.

Bagde, S., D. Epple, and L. Taylor (2016). Does affirmative action work? caste, gender, college quality, and academic success in India. *American Economic Review 106*(6), 1495–1521.

Bergman, P., E. Kopko, and J. E. Rodriguez (2021). Using predictive analytics to track students: Evidence from a seven-college experiment. NBER Working Paper No. 28948.

Bertrand, M., R. Hanna, and S. Mullainathan (2010). Affirmative action in education: Evidence from engineering college admissions in India. *Journal of Public Economics 94*(1), 16–29.

Bettinger, E. P., B. J. Evans, and D. G. Pope (2013). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy 5*(2), 26–52.

Black, S. E., J. T. Denning, and J. Rothstein (2023). Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes. *American Economic Journal: Applied Economics 15*(1), 26–67.

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after California's Proposition 209. *The Quarterly Journal of Economics 137*(1), 115–160.

Bleemer, Z. and A. Mehta (2021). College major restrictions and student stratification. Working Paper.

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.

Buitrago, M. G. and C. L. S. Blanco (2002). Nuevo examen de estado para ingreso a la educación superior: Logros y perspectivas. In *Elementos de política para la educación superior Colombiana*, pp. 79. Instituto Colombiano para el Fomento de la Educación Superior.

Bulman, G. (2015). The effect of access to college assessments on enrollment and attainment. *American Economic Journal: Applied Economics 7*(4), 1–36.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

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.

Card, D. and A. B. Krueger (1992). Does school quality matter? returns to education and the characteristics of public schools in the United States. *Journal of Political Economy 100*(1), 1–40.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2020). Income segregation and intergenerational mobility across colleges in the United States. *The Quarterly Journal of Economics 135*(3), 1567–1633.

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.

Cullen, J. B., M. C. Long, and R. Reback (2013). Jockeying for position: Strategic high school choice under Texas' Top Ten Percent plan. *Journal of Public Economics 97*, 32–48.

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.

Dalla Zuanna, A., K. Liu, and K. G. Salvanes (2022). Pulled-in and crowded-out: Heterogeneous outcomes of merit-based school choice. Centre for Economic Policy Research (CEPR) Discussion Paper DP16853.

Daugherty, L., P. Martorell, and I. McFarlin (2014). Percent plans, automatic admissions, and college outcomes. *IZA Journal of Labor Economics 3*(1), 1.

de Roux, N. and E. Riehl (2022). Disrupted academic careers: The returns to time off after high school. *Journal of Development Economics 156*, 102824.

Deming, D. J. and C. R. Walters (2017). The impact of price caps and spending cuts on US postsecondary attainment. National Bureau of Economic Research.

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 (2020). The consequences of academic match between students and colleges. *Journal of Human Resources 55*(3), 767–808.

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.

Durlauf, S. N. (2008). Affirmative action, meritocracy, and efficiency. *Politics, Philosophy & Economics 7*(2), 131–158.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income

students. *American Economic Review 111*(6), 1721–56.
El Tiempo (1999). Listos los cambios en la prueba ICFES. Feburary 7, 1999. Available at: https://www.eltiempo.com/archivo/documento/MAM-869108.
El Tiempo (2000). Nuevo ICFES revela fallas en educación. October 19, 2000. Available at: https://www.eltiempo.com/archivo/documento/MAM-1302187.
Goodman, J., O. Gurantz, and J. Smith (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy 12*(2), 115–58.
Goodman, J., M. Hurwitz, and J. Smith (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics 35*(3), 829–867.
Goodman, S. (2016). Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics 98*(4), 671–684.
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.
Hinrichs, P. (2014). Affirmative action bans and college graduation rates. *Economics of Education Review 42*, 43–52.
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. and S. Turner (2013). Expanding college opportunities for high-achieving, low income students. Stanford Institute for Economic Policy Research Discussion Paper.
Hoxby, C. M. (2009). The changing selectivity of American colleges. *The Journal of Economic Perspectives 23*(4), 95–118.
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.
Kain, J. F., D. M. O'Brien, and P. A. Jargowsky (2005). *Hopwood and the Top 10 Percent Law: How they have Affected the College Enrollment Decisions of Texas High School Graduates*. Texas School Project, University of Texas at Dallas.
Kapor, A. (2015). Distributional effects of race-blind affirmative action. Working Paper.
Kirkebøen, L., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.
Kobrin, J. L., B. F. Patterson, E. J. Shaw, K. D. Mattern, and S. M. Barbuti (2008). Validity of the SAT® for predicting first-year college grade point average. research report no. 2008-5. Technical report, College Board.
Lemann, N. (2000). *The big test: The secret history of the American meritocracy*. Macmillan.
Long, M. C. (2004). Race and college admissions: An alternative to affirmative action? *Review of Economics and Statistics 86*(4), 1020–1033.
Machado, C., G. Reyes, and E. Riehl (2022). The efficacy of large-scale affirmative action at elite universities. Working paper.

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.

Michelman, V., J. Price, and S. D. Zimmerman (2022). Old boys' clubs and upward mobility among the educational elite. *The Quarterly Journal of Economics 137*(2), 845–909.

Mountjoy, J. and B. Hickman (2020). The returns to college (s): Estimating value-added and match effects in higher education. University of Chicago, Becker Friedman Institute for Economics Working Paper.

Ng, K. and E. Riehl (2022). The returns to STEM programs for less-prepared students. Working Paper.

Niu, S. X. and M. Tienda (2010). The impact of the Texas Top Ten Percent Law on college enrollment: A regression discontinuity approach. *Journal of Policy Analysis and Management 29*(1), 84–110.

Ramey, G. and V. A. Ramey (2010). The rug rat race. *Brookings Papers on Economic Activity 41*(1 (Spring)), 129–199.

Riehl, E., J. E. Saavedra, and M. Urquiola (2019). Learning and earning: An approximation to college value added in two dimensions. In C. Hoxby and K. Stange (Eds.), *Productivity in Higher Education*, Chapter 4, pp. 105–132. The University of Chicago Press.

Riehl, E. and M. Welch (2023). Accountability, test prep incentives, and the design of math and English exams. *Journal of Policy Analysis and Management 42*(1), 60–96.

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics 121*(1), 297–317.

Saavedra, A. and J. E. Saavedra (2011). Do colleges cultivate critical thinking, problem solving, writing and interpersonal skills? *Economics of Education Review 30*(6), 1516–1526.

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.

Scott-Clayton, J. (2012). Do high-stakes placement exams predict college success? CCRC Working Paper No. 41.

Sekhri, S. (2020). Prestige matters: Wage premium and value addition in elite colleges. *American Economic Journal: Applied Economics 12*(3), 207–25.

Smith, J., J. Goodman, and M. Hurwitz (2020). The economic impact of access to public four-year colleges. National Bureau of Economic Research.

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 (2014). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies 81*(1), 426–472.

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.

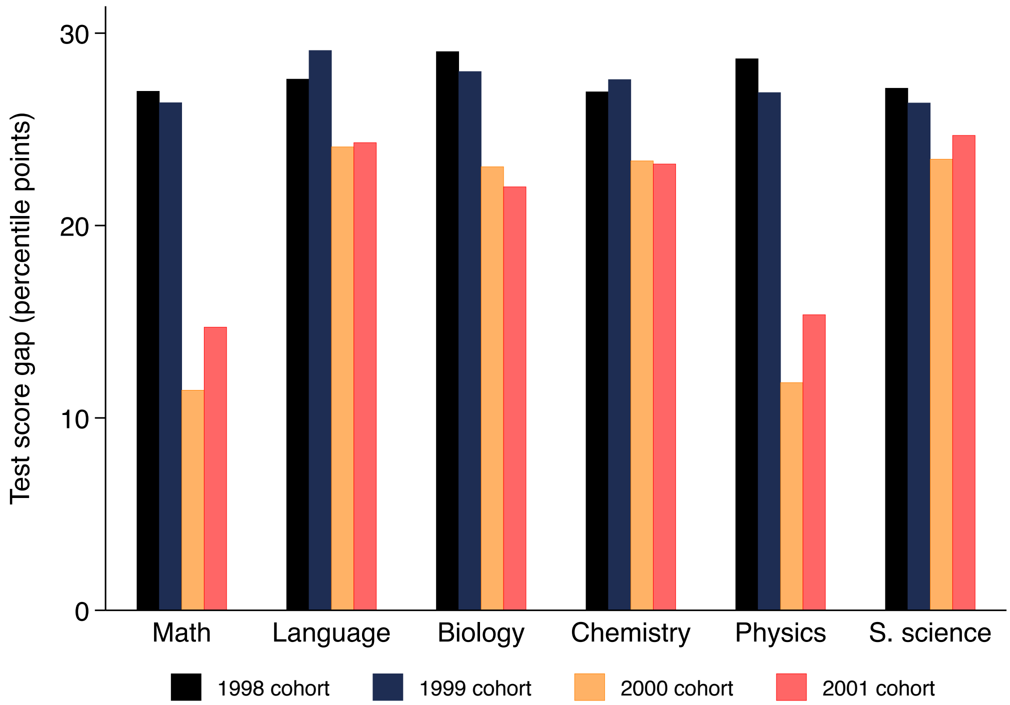
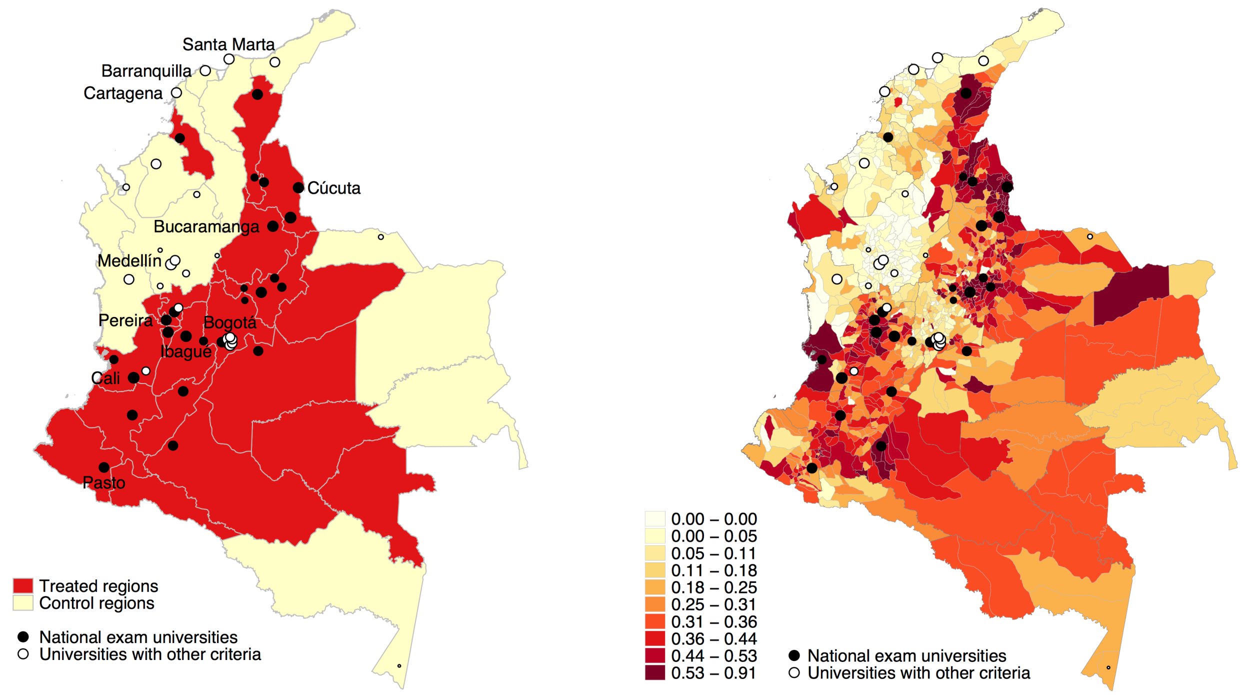


FIGURE 1. ICFES test score gaps by family income (*Top quartile* − *Bottom quartile*)

Notes: This graph shows gaps in ICFES scores between high- and low-income students. The y-axis is the test score gap by family income, defined as the difference in the average score percentile between students in the top and bottom income quartiles. The x-axis indicates the exam subject, and bar colors denote exam cohorts (1998–2001). See Appendix Table A2 for details on the subjects included in this figure and throughout the paper.



Panel A. Region-level treatment

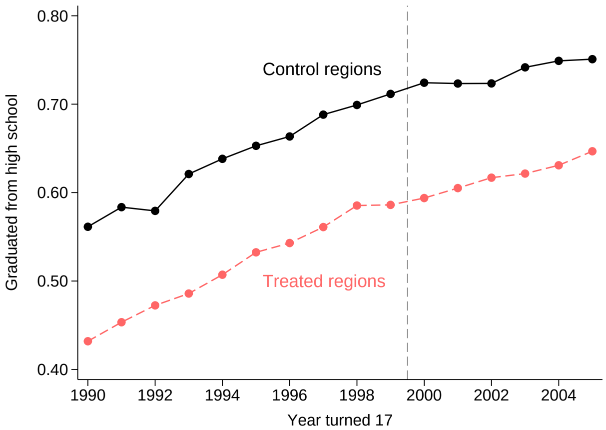
Panel B. Municipality-level treatment

FIGURE 2. Treatment variables

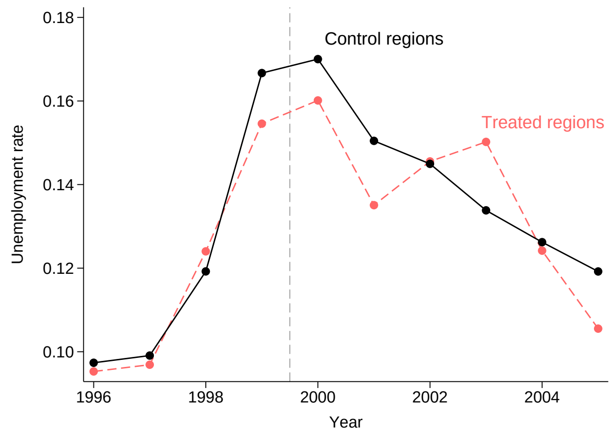
Notes: This figure displays my region- and municipality-level measures of exposure to ICFES exam universities. Black dots depict public universities that use only scores from the national ICFES exam in admissions. White dots are public universities with other admission criteria. Dot sizes are proportional to enrollment.

Panel A shows the region-level treatment variable, which is the (enrollment-weighted) modal public university admission method. I define treatment for regions with no universities using the closest public university to the capital city. Bogotá is its own administrative region, and the map does not show the island region San Andrés y Providencia. Both are control regions. See Appendix Table A3 for details.

Panel B defines treatment as the fraction of a municipality's college students from the 1998–1999 exam cohorts who enrolled in a public university with ICFES exam admissions. Colors depict deciles of this treatment variable, with darker shades reflecting higher treatment intensity.



Panel A. High school graduation rate



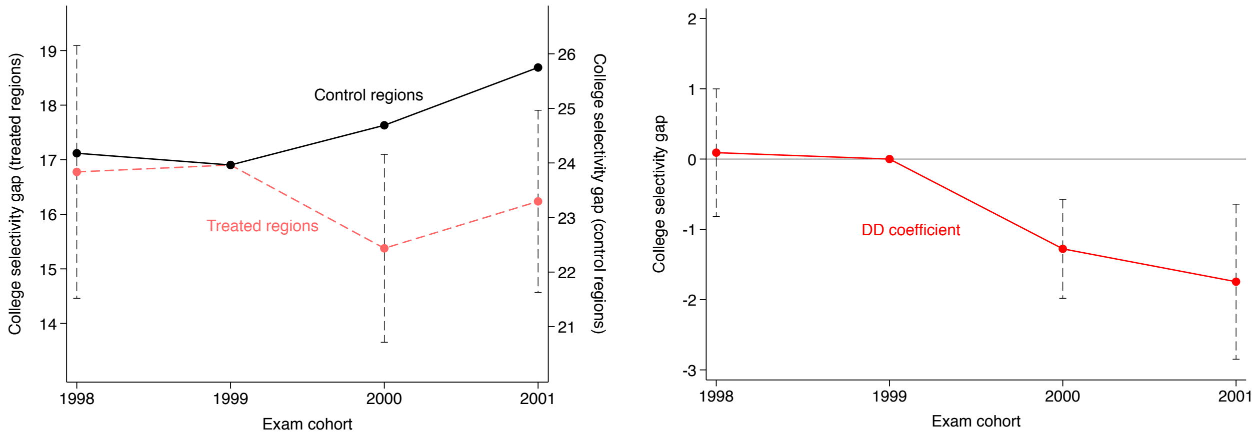
Panel B. Unemployment rate

Figure 3. High school graduation and unemployment rates in treated and control regions

Notes: This figure shows trends in educational attainment and macroeconomic conditions in treated and control regions.

Panel A shows the survey-weighted mean high school graduation rate in treated and control regions for cohorts defined by the year individuals turned 17. The vertical line separates cohorts who turned 17 before and after the admission exam reform. I use the 2011–2014 waves of the household survey *Gran Encuesta Integrada de Hogares*.

Panel B shows the mean unemployment rate in treated and control regions, with a vertical line separating pre- and post-reform years. Data are from the household survey *Encuesta Nacional de Hogares*. For 1996–2000, data are only available at the region level for the 24 largest regions, and the graph displays averages using population weights. For 2001–2005, I use individual level data for the same 24 regions and compute averages using survey weights.



Panel A. Treated and control regions

Panel B. Municipality-level event study

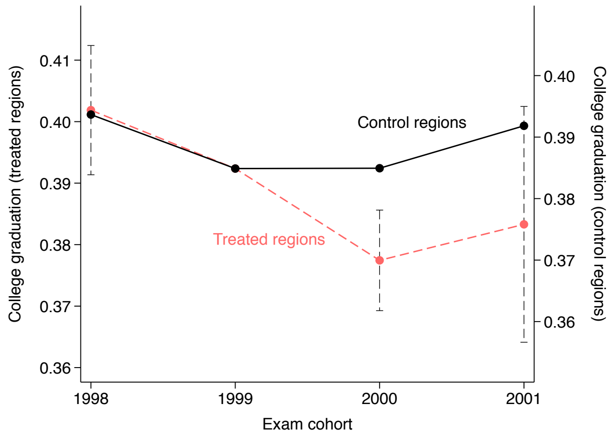
FIGURE 4. Reform effects on the college selectivity gap by family income
(*Top quartile* − *Bottom quartile*)

Notes: This figure shows the timing of the effects of the ICFES reform on college selectivity gaps by family income.

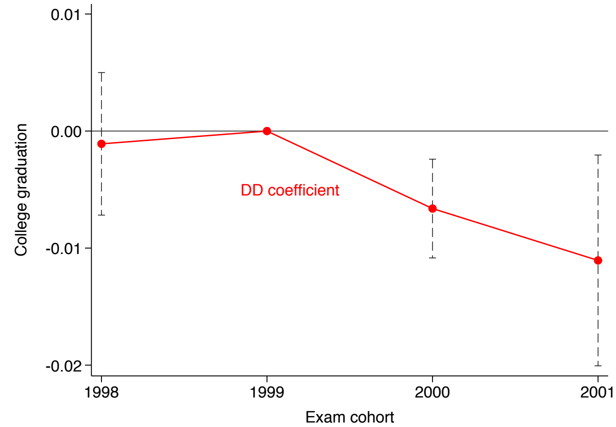
In Panel A, the dashed line is the difference in mean college selectivity between top and bottom income quartile students in treated regions. The solid line is the mean college selectivity gap in control regions.

Panel B plots event-study estimates of the reform's effect on the family income college selectivity gap using the municipality-level treatment variable and municipality fixed effects. The event study interacts the θ^q coefficient from equation (2) with dummies for exam cohorts t, omitting the 1999 cohort interaction.

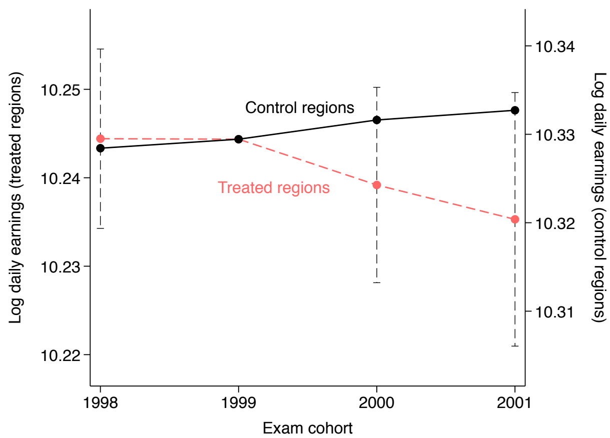
Dashed lines are 95 percent confidence intervals using standard errors from region- (Panel A) and municipality-level (Panel B) event-study regressions. Standard errors are clustered at the region level.



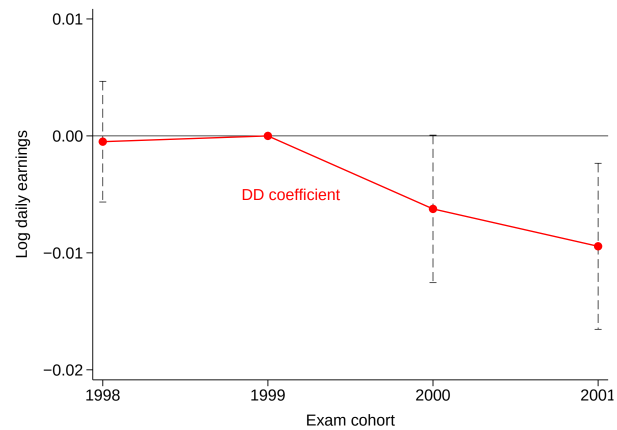
Panel A. College graduation
(*Treated and control regions*)



Panel B. College graduation
(*Municipality-level event study*)



Panel C. Log daily earnings
(*Treated and control regions*)



Panel D. Log daily earnings
(*Municipality-level event study*)

FIGURE 5. Reform effects on graduation and log daily earnings
(*All students*)

Notes: This figure illustrates the timing of the effects of the ICFES reform on college graduation (Panels A–B) and log daily earnings measured 10–11 years after the admission exam (Panels C–D). The sample includes all students in Column (B) of Table 1. Panels A and C plot means of each variable in treated regions (black solid lines) and control regions (red dashed lines). Panels B and D plot event-study estimates of the reform's effect on each variable using municipality-level treatment. The event study interacts the θ coefficient from equation (1) with dummies for cohorts t, omitting the 1999 cohort interaction.

Dashed lines are 95 percent confidence intervals using standard errors from region- (Panels A and C) and municipality-level (Panel B and D) event-study regressions. Standard errors are clustered at the region level.

Table 1. Summary statistics

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
	# students in 1998–2001 cohorts		Means for 1998–1999 cohorts				
			(in percentiles)			(in 2017 USD)	
SES group	High school graduates	College enrollees	ICFES math score	College selectivity	College grad	Daily earnings in 2008–12	Annual earnings in 2017
Top income quartile	223,417	152,358	70.4	62.9	0.48	23.75	10,892
Bottom income quartile	656,634	154,159	44.1	42.5	0.34	16.53	6,265
College educated mother	228,009	152,634	68.6	61.8	0.48	23.18	10,496
Primary educated mother	936,043	249,904	46.1	44.7	0.36	17.21	6,646
High ranked high school	315,699	200,771	69.3	63.1	0.48	22.90	10,272
Low ranked high school	841,124	203,277	42.9	38.9	0.31	16.34	6,139
All students	1,644,260	612,949	51.2	50.6	0.39	19.23	7,939

Notes: This table displays summary statistics for 1998–2001 ICFES exam takers. Each row reports totals/averages for the SES group listed in the first column. Column (A) shows the total number of students who took the ICFES exam in 1998–2001. Column (B) shows the number of these students who enrolled in any college in the Ministry of Education records. Columns (C)–(G) display means for the 1998–1999 exam cohorts, with column (C) including all exam takers and columns (D)–(G) including only college enrollees. Columns (C)–(D) report the mean ICFES math score and college selectivity in percentile units. Column (E) shows the fraction of college enrollees who graduated from their program. Column (F) shows mean daily earnings measured 10–11 years after individuals took the ICFES exam (converted to 2017 U.S. dollars). Column (G) shows mean annual earnings measured in 2017 (converted to 2017 U.S. dollars). See Appendix C.1 for details on the sample and variable definitions.

Table 2. Reform effects on test score gaps

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
		Dependent variable: ICFES exam score (percentile)					
		Reform effect by subject					
SES groups	Pre-reform mean gap	Math	Lang	Biol	Chem	Phys	S. sci
Family income gap	27.6	−13.7***	−4.2***	−6.0***	−4.0***	−14.3***	−2.7***
Top Q − Bottom Q		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Mother's education gap	24.2	−12.0***	−2.8***	−4.1***	−2.8***	−12.2***	−2.1***
College − Primary		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
High school rank gap	30.7	−17.1***	−3.2***	−5.2***	−4.0***	−17.7***	−1.3***
High − Low		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)

Notes: This table shows the effects of the ICFES exam reform on SES test score gaps. Column (A) shows mean test score gaps in percentile points across the six exam subjects for the 1998–1999 cohorts. (These gaps were similar in each exam subject.) For columns (B)–(G), I regress test scores in each subject on cohort dummies, an indicator for the high SES group, and the interaction of this indicator with a dummy for the 2000–2001 cohorts. Columns (B)–(G) display the coefficient on the interaction variable. The sample includes all ICFES exam takers in 1998–2001 (column A of Table 1). Parentheses contain robust standard errors.

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

Table 3. Reform effects on test validity

ICFES subject	(A) Exit exam score	(B) Graduated from college	(C) First year GPA	(D) S. science score
	Dependent variable (in st. deviation units)			
Panel A. Raw validity				
Math	−0.176***	−0.058***	−0.129**	−0.415***
	(0.004)	(0.003)	(0.061)	(0.001)
Language	−0.088***	0.009***	−0.003	−0.211***
	(0.004)	(0.003)	(0.052)	(0.001)
Biology	−0.039***	0.009***	−0.050	−0.258***
	(0.004)	(0.003)	(0.057)	(0.001)
Chemistry	0.013***	−0.003	−0.091	−0.169***
	(0.004)	(0.003)	(0.059)	(0.001)
Physics	−0.148***	−0.063***	−0.109*	−0.453***
	(0.004)	(0.003)	(0.059)	(0.001)
Social science	−0.008**	0.034***	−0.019	
	(0.004)	(0.003)	(0.063)	
Panel B. Within-SES validity				
Math	−0.186***	−0.055***	−0.116*	−0.358***
	(0.004)	(0.003)	(0.063)	(0.002)
Language	−0.105***	−0.001	0.017	−0.223***
	(0.004)	(0.003)	(0.053)	(0.001)
Biology	−0.056***	−0.002	−0.031	−0.267***
	(0.004)	(0.003)	(0.058)	(0.001)
Chemistry	−0.003	−0.017***	−0.050	−0.163***
	(0.004)	(0.003)	(0.058)	(0.001)
Physics	−0.161***	−0.060***	−0.078	−0.397***
	(0.004)	(0.003)	(0.060)	(0.001)
Social science	−0.021***	0.024***	−0.001	
	(0.004)	(0.003)	(0.064)	
N	242,887	612,949	3,084	1,644,260
Mean pre-reform correlation	0.386	0.112	0.093	0.714

Notes: This table shows how the ICFES reform affected the predictive power of each subject score. For Panel A, I regress the dependent variable on cohort dummies, the ICFES score, and the interaction of the ICFES score with a dummy for the 2000–2001 cohorts. The table displays coefficients on the interaction term. All covariates in column (C) regressions are interacted with dummies for the number of years since the admission exam. In columns (A)–(C), ICFES scores and dependent variables are residuals from regressing these variables on college × cohort fixed effects. In all columns, I then normalize variables to standard deviation one within each exam cohort so that coefficients can be interpreted as correlation coefficients. Panel B is identical to Panel A, except all ICFES scores and dependent variables are residualized on covariates for family income, mother's education, and high school rank.

The sample for column (A) includes all college enrollees who took the Colombian college *exit* exam. Column (B) includes all college enrollees (column B of Table 1). Column (C) includes students in the 2000–2004 entering cohorts at a public flagship university for which I have transcript data (see Appendix C.1). Column (D) includes all ICFES exam takers (column A of Table 1). The bottom row shows the mean (raw) correlation between the dependent variable and the six subject scores in the 1998–1999 cohorts. Parentheses contain robust standard errors.

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

Table 4. Treated and control region characteristics and balance tests

	(A)	(B)	(C)	(D)	(E)	(F)
	Pre-reform means		Region-level balance tests		Municipality-level balance tests	
Dependent variable	Treated regions	Control regions	Coef	SE	Coef	SE
Panel A. Student characteristics						
Top income quartile	0.227	0.279	−0.002	(0.003)	0.000	(0.001)
Bottom income quartile	0.264	0.235	0.001	(0.004)	−0.001	(0.002)
Mother's years of education	7.161	7.529	−0.087*	(0.045)	−0.035**	(0.016)
Father's years of education	7.455	7.970	−0.081	(0.062)	−0.033	(0.023)
Private high school	0.321	0.429	0.000	(0.009)	0.003	(0.003)
High ranked high school	0.181	0.209	−0.005	(0.005)	−0.001	(0.002)
Combined index (log earnings)	10.512	10.535	−0.000	(0.002)	0.001	(0.001)
Panel B. Labor market conditions						
Labor force participation rate	0.604	0.579	−0.001	(0.009)		
Unemployment rate	0.141	0.146	−0.006	(0.011)		
Panel C. Exam taking and college enrollment						
Annual exam takers (1000s)	11.545	16.474	0.508	(1.261)	0.014	(0.012)
Retook exam within one year	0.102	0.055	−0.010	(0.013)	−0.006	(0.005)
Enrolled in any college	0.343	0.362	−0.003	(0.009)	−0.000	(0.004)
Years between HS and college	3.567	3.372	−0.145	(0.087)	−0.059	(0.047)
Kilometers from HS to college	102.925	59.077	−3.648	(3.256)	−2.393	(1.561)
Stayed in region if enrolled	0.572	0.850	0.016*	(0.008)	0.009*	(0.005)
Public university if enrolled	0.468	0.275	0.030*	(0.016)	0.009	(0.006)
Regions	20	13				

Notes: The sample for Panel A includes all students in column (A) of Table 1. Income quartiles in Panel A are defined relative to all exam takers in each cohort. I compute the combined index by regressing 2012 log daily earnings on all covariates in Panel A using only pre-reform cohorts. The index is the predicted value from this regression in all cohorts. The data for Panel B are from the Colombian labor market survey *Encuesta Nacional de Hogares*. These data contain region-level statistics from the 24 largest regions. The sample for Panel C is the same as in Panel A, except the last four rows include only students who enrolled in college.

Columns (A)–(B) present means of each variable for 1998–1999 exam takers in treated and control regions. Columns (C)–(D) display estimates of θ and its standard error from equation (1) using the region-level treatment variable and region fixed effects. Columns (E)–(F) display analogous estimates using the municipality-level treatment variable and municipality fixed effects. Standard errors are clustered at the region level.

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

TABLE 5. Reform effects on SES college selectivity gaps

SES groups	(A) Benchmark regression	(B) Population controls	(C) Pop. & proximity controls
	Dependent variable: College selectivity, Q_c		
Panel A. Region-level treatment			
Family income gap	-2.36^{***}	-3.30^{***}	-3.47^{***}
Top Q − Bottom Q	(0.53)	(0.93)	(1.05)
Mother's education gap	-1.41^{***}	-1.18^{*}	-1.50^{*}
College − Primary	(0.49)	(0.67)	(0.76)
High school rank gap	−1.48	−1.09	−1.01
High − Low	(0.92)	(1.06)	(1.22)
Panel B. Municipality-level treatment			
Family income gap	-1.54^{***}	-1.60^{***}	-1.88^{***}
Top Q − Bottom Q	(0.36)	(0.46)	(0.56)
Mother's education gap	-0.80^{**}	−0.58	−0.76
College − Primary	(0.31)	(0.40)	(0.46)
High school rank gap	-1.31^{**}	-0.99^{*}	−0.97
High − Low	(0.48)	(0.52)	(0.65)

Notes: This table displays θ^q coefficients from separate regressions (2). The sample includes students in Column (B) of Table 1. Rows are defined by the three SES measures from Table 1. Panel A reports estimates using the region-level treatment variable (Treatment_g) and fixed effects for regions (γ_g and γ_{gt}). Panel B uses the municipality-level treatment variable and fixed effects for municipalities. In both panels, column (A) presents estimates from equation (2) as written. Column (B) adds dummies for eleven municipality population groups fully interacted with dummies for each exam cohort and SES group. Column (C) adds dummies for full interactions between municipality population groups, municipality proximity groups, exam cohorts, and SES groups. See the text and Appendix Table A5 for details on the population and proximity groups.

Parentheses contain standard errors clustered at the region level.

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

Table 6. Reform effects on college and labor market outcomes

SES group	(A) One year persistence	(B) College graduation	(C) Log daily earnings in 2008–12	(D) Log annual earnings in 2017
Panel A. Region-level treatment				
All students	-0.014^{***}	-0.015^{*}	-0.013^{**}	-0.024^{**}
	(0.005)	(0.008)	(0.005)	(0.009)
Top income quartile	-0.018^{*}	-0.027^{***}	-0.018^{*}	-0.042^{**}
	(0.009)	(0.009)	(0.009)	(0.019)
Bottom income quartile	-0.021^{***}	-0.010	-0.014^{***}	-0.018
	(0.006)	(0.009)	(0.005)	(0.013)
College educated mother	-0.016^{*}	-0.022^{**}	-0.011	-0.033
	(0.008)	(0.010)	(0.011)	(0.024)
Primary educated mother	-0.013^{**}	-0.007	-0.010^{**}	-0.020^{**}
	(0.005)	(0.008)	(0.005)	(0.008)
High ranked high school	-0.015^{**}	-0.019^{**}	-0.006	-0.024^{*}
	(0.007)	(0.007)	(0.009)	(0.013)
Low ranked high school	-0.014^{*}	-0.010	-0.013^{**}	-0.013
	(0.007)	(0.011)	(0.005)	(0.017)
Panel B. Municipality-level treatment				
All students	-0.010^{***}	-0.008^{**}	-0.007^{**}	-0.015^{***}
	(0.003)	(0.004)	(0.003)	(0.003)
Top income quartile	-0.012^{***}	-0.011^{**}	-0.008	-0.021^{**}
	(0.004)	(0.005)	(0.006)	(0.009)
Bottom income quartile	-0.013^{***}	-0.008^{*}	-0.008^{**}	-0.015
	(0.003)	(0.005)	(0.003)	(0.009)
College educated mother	-0.010^{**}	-0.008	-0.004	-0.021^{*}
	(0.004)	(0.006)	(0.006)	(0.011)
Primary educated mother	-0.010^{***}	-0.007	-0.007^{***}	-0.010^{*}
	(0.003)	(0.004)	(0.002)	(0.005)
High ranked high school	-0.010^{**}	-0.011^{**}	-0.006	-0.014^{**}
	(0.004)	(0.005)	(0.005)	(0.007)
Low ranked high school	-0.008^{**}	-0.004	-0.005	-0.009
	(0.003)	(0.005)	(0.003)	(0.007)
N (All students)	612,949	612,949	340,623	356,851

Notes: This table displays θ coefficients from separate regressions (1) with four dependent variables: an indicator equal to one if the student was still in college one year after enrolling (column A); an indicator for college graduation (column B); log daily earnings 10–11 years after the ICFES exam (column C); and log annual earnings in 2017 (column D). The sample includes students in Column (B) of Table 1 and the subset of those with formal earnings. The first row shows estimates for all students; other rows report estimates by SES group. Panel A reports estimates using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Panel B uses the municipality-level treatment variable and municipality dummies. Parentheses contain standard errors clustered at the region level.

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

Table 7. Heterogeneity in reform effects by field of study

Field/program group	(A) Prop. of students	(B) Pre-reform correlation w/ graduation: Math score	(C) Pre-reform correlation w/ graduation: Physics score	(D) Reform effects: Chose this field	(E) Reform effects: College grad	(F) Reform effects: Log daily earnings
Panel A. Field of program that student enrolled in						
STEM	0.349	0.074***	0.062***	0.002	−0.031***	−0.028**
		(0.005)	(0.005)	(0.006)	(0.008)	(0.010)
Business	0.271	0.075***	0.044***	0.003	−0.000	−0.016
		(0.007)	(0.007)	(0.006)	(0.010)	(0.009)
Health & humanities	0.380	0.032***	0.030***	−0.006	−0.009	0.001
		(0.004)	(0.003)	(0.005)	(0.010)	(0.008)
Panel B. Student's desired field at time of ICFES exam						
STEM	0.383	0.060***	0.043***	−0.004	−0.025**	−0.014
		(0.006)	(0.005)	(0.006)	(0.011)	(0.009)
Business	0.152	0.057***	0.043***	0.007	−0.017	−0.027**
		(0.007)	(0.009)	(0.005)	(0.012)	(0.010)
Health & humanities	0.466	0.042***	0.033***	−0.003	−0.004	−0.001
		(0.003)	(0.005)	(0.005)	(0.007)	(0.006)
Panel C. Importance of math and physics skill for graduating from program						
Most important quartile	0.252	0.109***	0.090***	0.010	−0.032***	−0.017**
		(0.004)	(0.006)	(0.007)	(0.006)	(0.008)
Quartile 3	0.252	0.040***	0.047***	−0.005	−0.013	−0.021**
		(0.007)	(0.005)	(0.011)	(0.011)	(0.009)
Quartile 2	0.248	0.009	0.013*	0.000	−0.016	−0.005
		(0.006)	(0.007)	(0.008)	(0.012)	(0.008)
Least important quartile	0.249	−0.013**	−0.038***	−0.006	−0.003	0.004
		(0.006)	(0.005)	(0.009)	(0.011)	(0.009)

Notes: This table examines heterogeneity in the effects of the ICFES reform. Panel A explores heterogeneity based on the field of study of students' college programs. I define three field groups based on the Ministry of Education's classification of programs into eight areas: 1) STEM (engineering and natural sciences); 2) Business; and 3) Health & humanities (health, fine arts, education, social sciences, and agronomy). Panel B explores heterogeneity using the same groups, but fields are based on students' desired programs as reported in the ICFES exam data. Panel C explores heterogeneity based on the predictive power of math and physics scores for graduation in each program. For each program, I regress an indicator for college graduation on all six ICFES subject scores (math, physics, language, biology, chemistry, social science) in the 1998–1999 cohorts. I then group programs into quartiles based on the sum of the math and physics coefficients from this regression.

Column (A) shows the proportion of students who chose each field/program in the pre-reform cohorts. In columns (B)–(C), I use the same regression as in Panel C, but I estimate it separately for each field/program group. I report the math (column B) and physics (column C) coefficients from these regressions in SD units. Columns (D)–(F) show estimates of θ from equation (1). In column (D), the sample includes all college enrollees in the 1998–2001 cohorts, and the dependent variable is an indicator for the field/program group in the row header. In columns (E)–(F), the samples include only students who chose the program/field group, and the dependent variables are an indicator for college graduation (column E) and log daily earnings 10–11 years after the ICFES exam (column F).

Parentheses contain standard errors clustered at the region level.

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

Appendix — For Online Publication

Outline:

A. Appendix figures and tables

B. Theoretical appendix

C. Empirical appendix

A. Appendix figures and tables

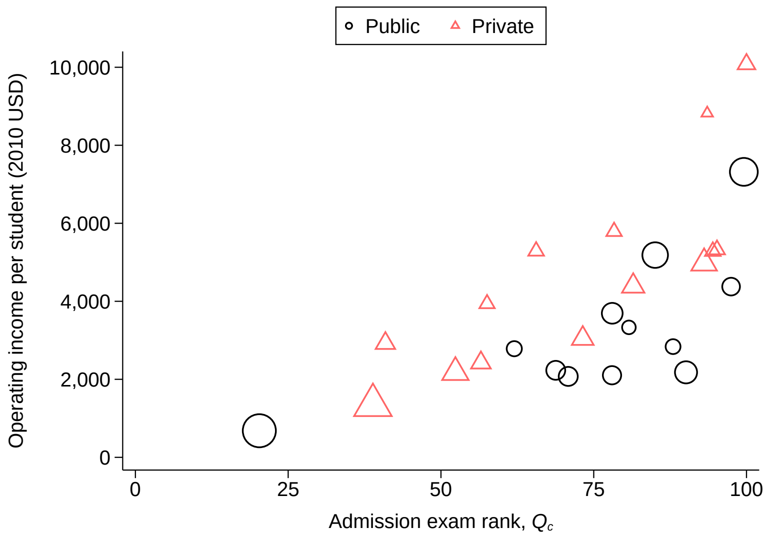
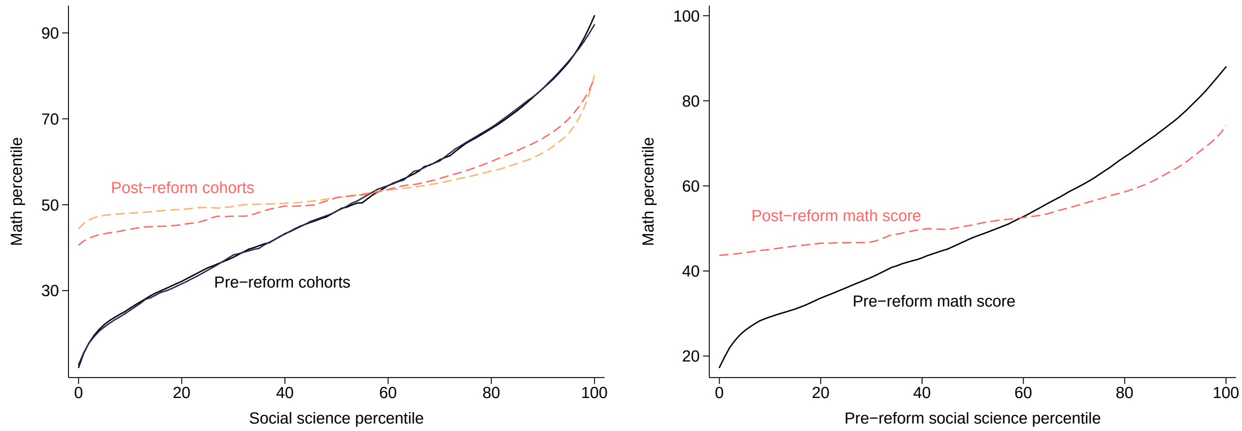


Figure A1. Admission exam rank, Q_c and operating income per student for 26 colleges

Notes: The x-axis is my main college selectivity measure Q_c. The y-axis depicts operating income per student in 2010 U.S. dollars. Income data are from a Ministry of Education list of the top 26 colleges by 2010 operating income (available in May 2018 at: https://www.mineducacion.gov.co/observatorio/1722/article-280538.html). I compute the number of students at each university for the year 2013 from the Ministry of Education's administrative enrollment records used elsewhere in this paper. The enrollment-weighted correlation between these two variables is 0.77. The unweighted correlated is 0.66.

Black circles are public colleges; red triangles are private colleges. Marker sizes are proportional to the number of students.

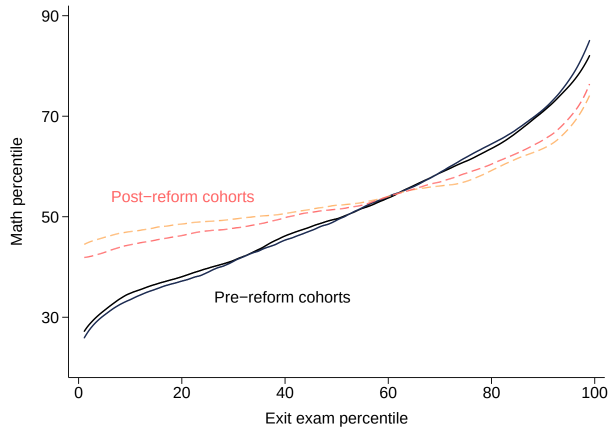


Panel A. All admission exam takers

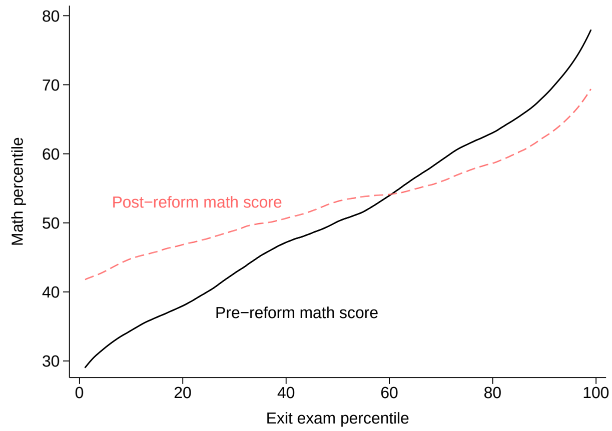
Panel B. Students who took *both* the pre- and post-reform exams

Figure A2. Relationship between math and social science admission scores

Notes: This figure shows the relationship between individuals' math and social science scores in the pre- and post-reform ICFES exams. Panel A plots local linear regressions of admission exam math percentiles on social science percentiles in a sample that includes all ICFES exam takers in the 1998–2001 cohorts (column A of Table 1). Panel B plots analogous regressions for the subset of students who took *both* the pre- and post-reform admission exams ($N = 40{,}063$). This panel plots estimates separately for pre- and post-reform math percentiles using pre-reform social science percentiles in all regressions. Percentiles are computed relative to the sample for each panel.



Panel A. All exit exam takers



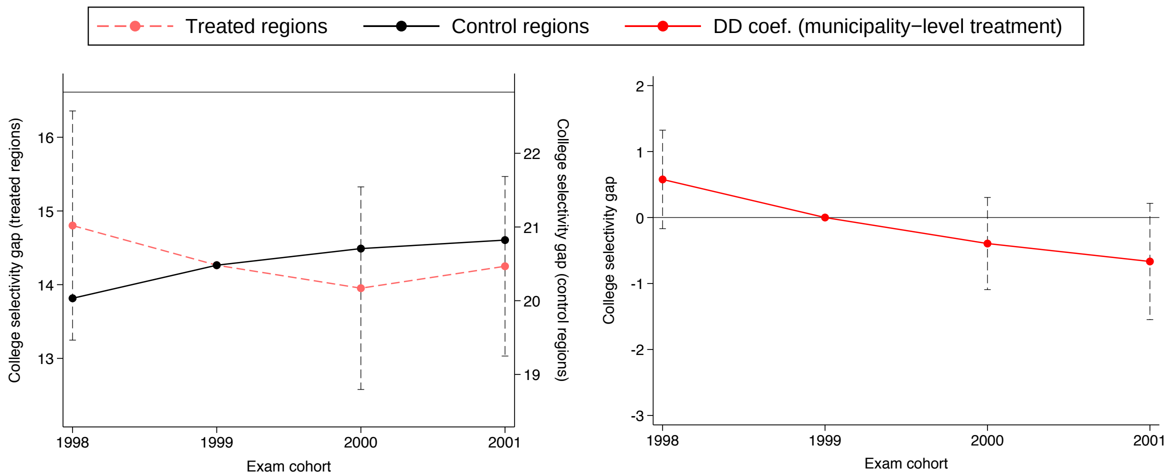
Panel B. Exit exam takers who took *both* admission exams

Figure A3. Relationship between math admission scores and exit exam scores

Notes: This figure is analogous to Figure A2, but the it uses scores on a field-specific college exit exam—rather than social science admission scores—as a measure of individual ability. Exit exam scores are not a perfect measure because they are confounded by the admission exam's effects on students' college choices. Nonetheless, the dramatic reductions in validity suggest that selection issues are likely to be a second order effect relative to changes in the abilities measured by the exams.

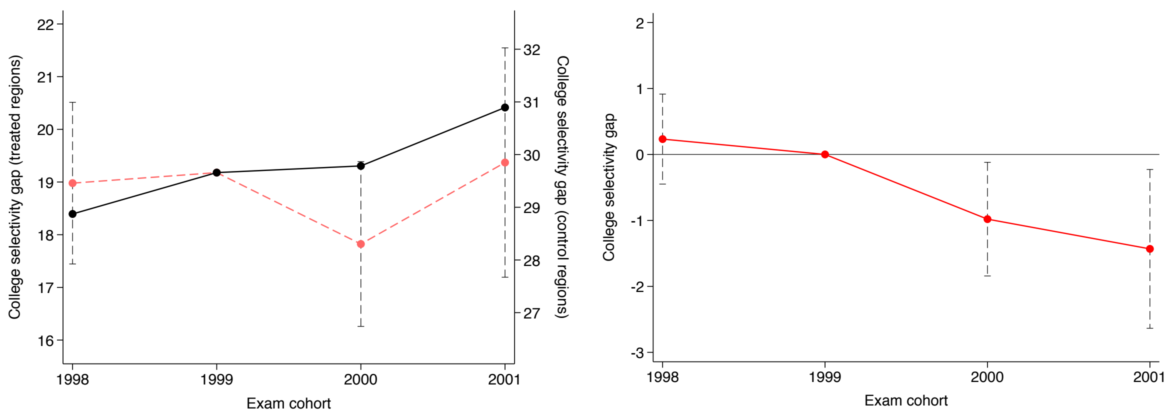
Panel A plots local linear regressions of admission exam math percentiles on exit exam percentiles for each cohort. The sample includes all students who took the exit exam ($N = 281{,}473$). Percentiles are computed relative to students in this subsample.

Panel B plots analogous regressions for the subset of exit exam takers who took *both* the pre- and post-reform admission exams ($N = 10{,}370$). I plot estimates separately for pre- and post-reform math percentiles. Percentiles are computed relative to students in this subsample.



Panel A. Mother's education gap
(*Treated and control regions*)

Panel B. Mother's education gap
(*Municipality-level event study*)

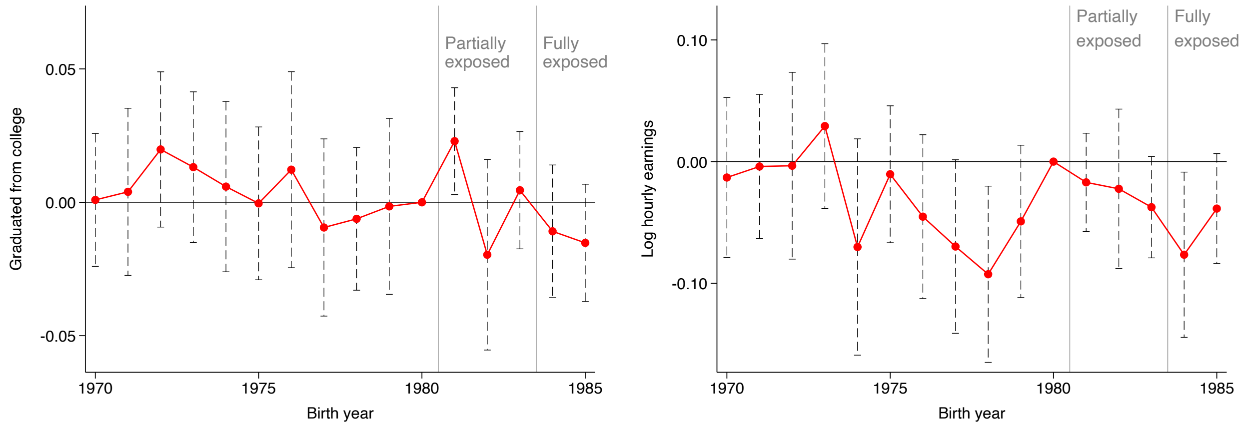


Panel C. High school rank gap
(*Treated and control regions*)

Panel D. High school rank gap
(*Municipality-level event study*)

FIGURE A4. Reform effects on SES college selectivity gaps

Notes: This figure plots region- and municipality-level estimates of the reform's effect on SES college selectivity gaps using the mother's education and high school rank measures of socioeconomic inequality defined in Table 1. Panels A and C are analogous to Panel A in Figure 4. Panels B and D are analogous to Panel B in Figure 4. See the notes to Figure 4 for details.



Panel A. Graduated from college

Panel B. Log hourly wage

FIGURE A5. Event-study estimates for college graduation and earnings in household survey data

Notes: This figure displays event-study estimates of the effects of the 2000 ICFES reform on region-level mean college graduation rates and log hourly earnings using Colombian household survey data. The data are from the January 2015 through June 2019 waves of the household survey *Gran Encuesta Integrada de Hogares*.

In Panel A, the dependent variable is an indicator for completing a technical college, university, or post-graduate degree. In Panel B, the dependent variable is log hourly earnings, defined as log(Last month's earnings/(Usual weekly hours $\times$ 30/7)). The sample for both panels includes all surveyed individuals who were born between 1970–1985 and who enrolled in a technical college or university. Panel B additionally restricts to individuals with a non-zero hourly wage.

The figures display θ_c coefficients from the event-study regression:

$$y_{icrt} = \gamma_c + \gamma_r + \gamma_{a(ct)} + \theta_c \text{Treated}_r + u_{icrt}$$

where y_{icrt} is an outcome for individual i who was born in year c, lives in region r, and was surveyed in year t. The regression includes birth year fixed effects, γ_c, region fixed effects, γ_r, and age fixed effects, $\gamma_{a(ct)}$, where $a(ct) = t - c$. The coefficients of interest, θ_c, are on the interaction between a dummy for treated regions, Treated$_r$, and indicators for each birth year (omitting 1980). Since most individuals take the ICFES exam between the ages of 16–19, the graphs denote groups of birth cohorts that were not exposed (1970–1980), partially exposed (1981–1983), and fully exposed (1984–1985) to the 2000 ICFES exam reform. All regressions are estimated using survey weights. Vertical dashed lines are 95 percent confidence intervals using standard errors clustered at the region level.

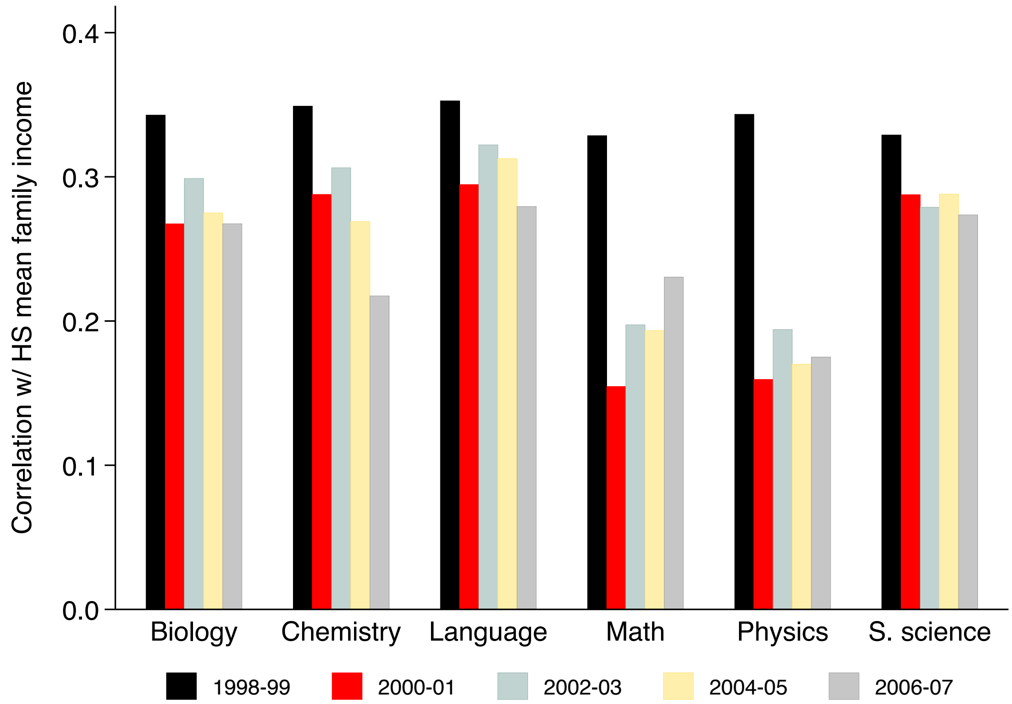


FIGURE A6. Correlation between ICFES test scores and HS mean family income

Notes: This graph shows correlations between ICFES scores and mean family income at the high school level. I first compute family income in the pre-reform cohorts (1998–1999) using the same method as described in Section 2.2. I then compute the average family income in each high school and use this average for all cohorts from 1998–2007. The y-axis shows the correlation coefficient between this value and scores on the ICFES exam subject listed on the x-axis. Bar colors denote exam cohort groups from 1998–2007. See Appendix Table A2 for details on the subjects included in this figure and throughout the paper.

Table A1. Admission exam rank, Q_c, and other college characteristics by institution tier

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Institutional level	College ownership	Total enrollment	Mean years to grad	Admission exam rank (Q_c)	Admit rate	Grad rate	Annual tuition (2012 USD)
University	Private	484,487	5.6	55.0	0.82	0.45	2,877
University	Public	543,064	5.6	58.7	0.47	0.42	440
University Institute	Private	300,852	4.4	29.5	0.86	0.39	1,270
University Institute	Public	95,063	4.0	19.0	0.72	0.36	472
Technology Institute	Private	51,307	3.1	29.1	0.90	0.26	950
Technology Institute	Public	35,242	3.5	46.9	0.85	0.39	799
Technical/Professional Institute	Private	40,173	3.1	22.7	0.83	0.24	830
Technical/Professional Institute	Public	6,787	2.8	5.5	0.93	0.39	415

Notes: This table shows how my college selectivity measure, Q_c, relates to other institutional characteristics.

Columns (A) and (B) categorize colleges into tiers based on their institution level and ownership status. Technology Institutes also include the Ministry of Education's fifth category of college, Technology Schools.

Column (C) shows total enrollment in each tier for the year 2013. Column (D) shows average time to graduation. Column (E) shows the average value of my main college selectivity measure, Q_c. Column (G) shows the average graduation rate in the 2002–2003 enrollment cohorts. Each of these statistics is computed from the Ministry of Education's administrative enrollment records used elsewhere in this paper.

Column (F) shows the average admission rate (number admitted/number applied) over the years 2007–2013 using college-program-cohort level data from the Ministry of Education (available in May 2018 at: https://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html). Column (H) shows the median annual tuition rate in each tier for 2012 as reported by the Ministry of Education (available in May 2018 at: https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-212350_resumen.xls). Within each institution tier, I compute a weighted average of the median tuitions for each of three "program levels" (technical professional, technical, and university) using the number of students who enrolled in each program level as weights.

Table A2. ICFES exam summary statistics by subject test and cohort

(A)	(B)	(C)				(D)				(E)			
		Mean score				SD of scores				# unique scores			
Subject	Tests	1998	1999	2000	2001	1998	1999	2000	2001	1998	1999	2000	2001
Biology	Biology	48.0	48.3	45.1	44.6	10.0	10.1	5.8	5.4	56	51	39	36
Chemistry	Chemistry	45.7	50.3	44.9	45.0	9.7	10.1	5.7	5.6	46	41	39	38
Language	Language	48.8	50.8	46.4	46.4	10.1	10.0	6.4	6.0	56	60	36	36
Math	Math aptitude	49.2	50.5	.	.	9.9	9.9	.	.	32	31	.	.
	Math knowledge	48.8	48.6	.	.	9.8	9.9	.	.	38	31	.	.
	Math	.	.	43.1	41.2	.	.	5.5	5.4	.	.	34	40
Physics	Physics	47.1	46.7	45.2	46.7	9.7	9.9	6.0	6.1	50	41	32	33
Soc. science	Social sciences	48.2	48.6	.	.	10.2	8.5	.	.	70	61	.	.
	Geography	.	.	44.6	43.5	.	.	6.2	7.0	.	.	36	37
	History	.	.	43.7	43.5	.	.	6.1	6.4	.	.	38	37
Excluded components	Verbal aptitude	48.6	.	.	.	9.8	.	.	.	62	.	.	.
	Philosophy	.	.	44.8	43.7	.	.	6.3	6.2	.	.	38	38
	Foreign language	.	.	40.9	42.0	.	.	6.8	7.2	.	.	54	56
	Elective	49.8	50.9	52.2	55.4	11.4	11.8	10.8	8.8	70	65	49	42
	All tests	48.2	49.3	45.1	45.2	10.1	10.0	6.6	6.4	53	48	40	39

Notes: This figure displays summary statistics for ICFES scale scores by subject test and exam cohort. The sample includes all exam takers in the ICFES records. Column (A) shows the subject groups used in the paper, and column (B) shows the subject tests that comprise each group. Block (C) show mean scale scores by exam cohort. Block (D) show the standard deviation of scale scores by cohort. Block (E) show the number of unique score values averaged across the spring and fall administrations of each cohort.

The paper focuses on the first six subjects in this table: biology, chemistry, language, math, physics, and social science. Through the paper, math scores for the 1998–1999 cohorts average over the aptitude and knowledge components, and social science scores for the 2000–2001 cohorts average over the geography and history components.

Table A3. Public universities and definition of treated regions

(A)	(B)	(C)	(D)	(E)	(F)
Region	Public university	Cohort size	Admission method	Closest region	Treated region?
Amazonas	Univ. Nacional de Colombia (Leticia)	28	Other		
Antioquia	Univ. de Antioquia (Medellin)	6,962	Other		
	Univ. Nacional de Colombia (Medellin)	2,253	Other		
	Univ. de Antioquia (Carmen De Viboral)	220	Other		
	Univ. de Antioquia (Turbo)	165	Other		
	Univ. de Antioquia (Caucasia)	128	Other		
	Univ. de Antioquia (Andes)	101	Other		
	Univ. de Antioquia (Puerto Berrio)	50	Other		
	Univ. de Antioquia (Antioquia)	49	Other		
Arauca	Univ. Nacional de Colombia (Arauca)	65	Other		
Atlantico	Univ. del Atlantico	3,609	Other		
Bogota	Univ. Nacional Abierta y A Distancia	14,589	Other		
	Univ. Distrital Francisco Jose de Caldas	5,095	National exam		
	Univ. Nacional de Colombia (Bogota)	4,882	Other		
	Univ. Ped. Nacional	1,910	Other		
	Univ. Militarnueva Granada	1,896	Other		
	Univ. Colegio Mayor de Cundinamarca	1,365	Other		
Bolivar	Univ. de Cartagena	2,978	Other		
Boyaca	Univ. Ped. y Tecn. de Colombia (Tunja)	3,196	National exam		✓
	Univ. Ped. y Tecn. de Colombia (Sogamoso)	548	National exam		
	Univ. Ped. y Tecn. de Colombia (Duitama)	490	National exam		
	Univ. Ped. y Tecn. de Colombia (Chiquinquira)	150	National exam		
Caldas	Univ. de Caldas	2,810	National exam		✓
	Univ. Nacional de Colombia (Manizales)	991	Other		
Caqueta	Univ. de la Amazonia	1,274	National exam		✓
Casanare		431		Boyaca	✓
Cauca	Univ. del Cauca	2,221	National exam		✓
Cesar	Univ. Popular del Cesar (Valledupar)	2,488	National exam		✓
	Univ. Popular del Cesar (Aguachica)	233	National exam		
Choco	Univ. Tecn. del Chocodiego Luis Cordoba	2,278	Other		
Cordoba	Univ. de Cordoba	2,210	Other		
Cundinamarca	Univ. de Cundinamarca (Fusagasuga)	2,013	National exam		✓
	Univ. de Cundinamarca (Girardot)	421	National exam		
	Univ. de Cundinamarca (Ubate)	134	National exam		
Guainia		8		Arauca	
Guaviare		58		Meta	✓
Huila	Univ. Surcolombiana	1,750	National exam		✓

Notes: This table is continued on the next page.

Table A3. Public universities and definition of treated regions (continued)

(A) Region	(B) Public university	(C) Cohort size	(D) Admission method	(E) Closest region	(F) Treated region?
La Guajira	Univ. de la Guajira	1,970	Other		
Magdalena	Univ. del Magdalena	3,257	Other		
Meta	Univ. de Los Llanos	1,148	National exam		✓
Narino	Univ. de Narino	1,996	National exam		✓
Norte Santander	Univ. de Pamplona	6,192	National exam		✓
	Univ. Francisco de Paula Santander (Cucuta)	3,387	National exam		
	Univ. Francisco de Paula Santander (Ocana)	944	National exam		
Putumayo		260		Narino	✓
Quindio	Univ. del Quindio	3,118	National exam		✓
Risaralda	Univ. Tecn. de Pereira	2,701	National exam		✓
San Andres		102		Bolivar	
Santander	Univ. Industrial de Santander	3,838	National exam		✓
Sucre	Univ. de Sucre	1,033	National exam		✓
Tolima	Univ. del Tolima	4,099	National exam		✓
Valle del Cauca	Univ. del Valle	5,660	National exam		✓
	Univ. del Pacifico	539	National exam		
	Univ. Nacional de Colombia (Palmira)	517	Other		
Vaupes		12		Meta	✓
Vichada		27		Arauca	

Notes: This table describes my definition of treated and control regions, which is based on pre-reform admission methods in public universities.

Column (A) lists the 33 Colombian administrative regions (*departamentos*).

Column (B) lists public universities in each region; blank cells indicate that the region has no public universities.

Column (C) shows the mean number of entering students per cohort over all cohorts in my records. For regions without public universities, it shows the average number of public university enrollees per pre-reform exam cohort.

Column (D) lists the pre-reform admission method of each university. I collected information on admission methods prior to 2000 by searching through historical student regulations at each college, or by tracking down information from historical college or newspaper websites using the website archive.org. "National exam" means that the university used only ICFES national exam scores for admission. "Other" means that the university used other admission criteria; most commonly this meant that applicants were required to take the university's own admission exam, but in some cases schools considered other information such as high school GPA or personal interviews.

Column (E) shows the closest region with a public university for those regions with blank cells in column (B). Closest is defined by distance between capital cities.

Column (F) shows my classification of regions as treated or control. Treated regions are those where public universities used only the national ICFES exam for admissions. Control regions are those with public universities that use other admission methods. In the three regions with mixed admission criteria (Bogotá, Caldas, and Valle del Cauca), I define treatment as the modal admission method (weighted by column C). For regions with no public universities, I define treatment using the region listed in column (E).

Table A4. Reform effects on exam retaking, college enrollment, and formal sector employment

	(A)	(B)	(C)	(D)	(E)	(F)
	Dep. variable: Retook exam within one year		Dep. variable: Enrolled in any college		Dep. variable: Has formal earnings 10–11 years later	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	-0.010	-0.006	-0.003	-0.000	-0.010	-0.008^{**}
	(0.013)	(0.005)	(0.009)	(0.004)	(0.010)	(0.004)
Top income quartile	-0.013	-0.005	-0.022	-0.008	-0.020	-0.014^{**}
	(0.013)	(0.006)	(0.015)	(0.008)	(0.016)	(0.006)
Bottom income quartile	-0.005	-0.004	0.001	0.002	-0.004	-0.006
	(0.011)	(0.004)	(0.007)	(0.003)	(0.009)	(0.005)
College educated mother	-0.010	-0.004	-0.001	-0.001	-0.017	-0.014^{*}
	(0.014)	(0.006)	(0.012)	(0.007)	(0.017)	(0.007)
Primary educated mother	-0.009	-0.006	0.003	0.002	-0.002	-0.004
	(0.011)	(0.004)	(0.008)	(0.004)	(0.007)	(0.004)
High ranked high school	0.004	0.001	-0.008	-0.007	-0.014	-0.011
	(0.014)	(0.007)	(0.012)	(0.007)	(0.018)	(0.007)
Low ranked high school	-0.017^{**}	-0.008^{***}	0.004	0.005	-0.011	-0.007^{**}
	(0.007)	(0.003)	(0.008)	(0.004)	(0.008)	(0.003)
N (All students)	1,297,104	1,297,104	1,644,260	1,644,260	612,949	612,949

Notes: This table displays θ coefficients from separate regressions (1). The dependent variable in columns (A)–(B) is an indicator equal to one if the student retook the national exam within one year. The dependent variable in columns (C)–(D) is an indictor for enrolling in any college in the Ministry of Education records. The dependent variable in columns (E)–(F) is an indicator for appearing in the formal earnings records 10–11 years after the ICFES exam.

The sample for columns (C)–(D) includes all students in the 1998–2001 exam cohorts, i.e., those in column (A) of Table 1. The sample for columns (A)–(B) includes only the 1998–2000 exam cohorts because I do not observe exam retaking beyond 2001. The sample for columns (E)–(F) includes all college enrollees in the 1998–2001 cohorts (column B of Table 1). The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A), (C), and (E) report estimates using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Columns (B), (D), and (F) use the municipality-level treatment variable and municipality dummies.

Parentheses contain standard errors clustered at the region level.

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

Table A5. Municipality population and proximity groups for robustness tests

	(A) Exam takers per cohort	(B) Km from public uni	(C) Municipalities
1	26,831–79,592	0	Bogota, *Cali*, Medellin
2	5,054–17,045	0	Barranquilla, *Bucaramanga*, Cartagena, *Cucuta*, *Ibague*, *Manizales*, Monteria, *Neiva*, *Pasto*, *Pereira*, *Popayan*, Santa Marta, *Villavicencio*
3	1,633–4,378	0	*Armenia*, *Buenaventura*, *Duitama*, *Florencia*, *Girardot*, *Palmira*, Quibdo, Riohacha, *Sincelejo*, *Sogamoso*, *Tunja*, *Valledupar*
4	1,633–4,378	6–14	Bello, *Dos Quebradas*, Envigado, *Floridablanca*, Itagui, Soledad
5	1,633–4,378	20–32	Cerete, *Soacha*
6	1,633–4,378	32–40	*Buga*
7	1,633–4,378	53–85	*Barrancabermeja*, *Tulua*
8	690–1,504	0	*Aguachica*, Arauca, *Chiquinquira*, *Fusagasuga*, *Ocana*, *Pamplona*
9	690–1,504	6–14	*Calarca*, Copacabana
10	690–1,504	14–20	Caldas, *Corozal*, *Jamundi*, *Piedecuesta*, *Santa Rosa de Cabal*, *Yumbo*
11	690–1,504	20–32	Apartado, Baranoa, *Cartago*, *Espinal*, *Florida*, *Giron*, Rionegro
12	690–1,504	40–53	Cienaga, Fundacion, Lorica, *Madrid*, *Mosquera*, Sabanalarga, Sahagun, *Santander de Quilichao*, *Zipaquira*
13	690–1,504	53–85	*Chia*, *Facatativa*, *Ipiales*, *La Dorada*, Magangue, Maicao, *Pitalito*, *San Gil*, *Yopal*
14	690–1,504	141–231	*Tumaco*
15–23	282–660	9 groups	0km (*1T*, 5C), 6–14km (*5T*, 3C), 14–20km (*4T*, 7C), 20–32km (*9T*, 4C), 32–40km (*6T*, 7C), 40–53km (*13T*, 6C), 53–85km (*15T*, 10C), 86–134km (*2T*, 5C), 274–719km (*0T*, 1C)
24–32	174–272	9 groups	0km (*0T*, 2C), 6–14km (*3T*, 1C), 14–20km (*4T*, 1C), 20–32km (*6T*, 7C), 32–40km (*8T*, 3C), 40–53km (*19T*, 7C), 53–85km (*15T*, 5C), 86–134km (*5T*, 3C), 141–231km (*2T*, 1C),
33–40	85–172	8 groups	6–14km (*8T*, 3C), 14–20km (*16T*, 4C), 20–32km (*34T*, 11C), 32–40km (*19T*, 8C), 40–53km (*39T*, 13C), 53–85km (*35T*, 14C), 86–134km (*6T*, 3C), 141–231km (*0T*, 1C)
41–49	54–84	9 groups	6–14km (*4T*, 0C), 14–20km (*7T*, 2C), 20–32km (*14T*, 10C), 32–40km (*16T*, 6C), 40–53km (*13T*, 5C), 53–85km (*33T*, 16C), 86–134km (*6T*, 6C), 141–231km (*1T*, 0C)274–719km (*1T*, 2C),
50–58	26–54	9 groups	6–14km (*8T*, 3C), 14–20km (*9T*, 3C), 20–32km (*23T*, 6C), 32–40km (*27T*, 3C), 40–53km (*20T*, 11C), 53–85km (*49T*, 18C), 86–134km (*9T*, 8C), 141–231km (*6T*, 0C), 274–719km (*0T*, 1C)
59–67	12–26	9 groups	6–14km (*12T*, 1C), 14–20km (*13T*, 0C), 20–32km (*24T*, 2C), 32–40km (*16T*, 3C), 40–53km (*25T*, 5C), 53–85km (*36T*, 10C), 86–134km (*8T*, 6C), 141–231km (*6T*, 1C), 274–719km (*1T*, 1C),
68–76	2–11	9 groups	6–14km (*4T*, 0C), 14–20km (*2T*, 1C), 20–32km (*6T*, 1C), 32–40km (*3T*, 0C), 40–53km (*6T*, 1C), 53–85km (*22T*, 1C), 86–134km (*6T*, 4C), 141–231km (*1T*, 2C), 274–719km (*1T*, 0C)

Notes: Column (A) uses Ward's method for hierarchical clustering to divide municipalities into ten groups by the number of national exam takers per cohort. I create a separate eleventh group for Bogotá, Cali, and Medellín.

Column (B) uses Ward's method to divide municipalities into ten groups based on distance (in kilometers) from the nearest public university. Distance is defined as the crow flies using the capital city of each municipality.

Regressions in columns (C)–(D) of Table 5 include fully interacted dummies for population group/exam cohort/SES. Columns (E)–(F) in Table 5 include dummies for population group/proximity group/exam cohort/SES.

Italicized municipalities are in treated regions (T); non-italicized municipalities are in control regions (C).

Table A6. Reform effects on one year college persistence

	(A)	(B)	(C)	(D)	(E)	(F)
	Dependent variable: One year college persistence					
	Benchmark specification (1)		Within population groups		Within population × proximity groups	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	-0.0142^{***}	-0.0096^{***}	-0.0092	-0.0084^{**}	-0.0073	-0.0076^{*}
	(0.0051)	(0.0028)	(0.0071)	(0.0036)	(0.0075)	(0.0042)
Top income quartile	-0.0176^{*}	-0.0120^{***}	-0.0071	-0.0089	-0.0044	-0.0089
	(0.0091)	(0.0042)	(0.0142)	(0.0061)	(0.0147)	(0.0070)
Bottom income quartile	-0.0208^{***}	-0.0128^{***}	-0.0180^{**}	-0.0118^{***}	-0.0148^{**}	-0.0102^{**}
	(0.0057)	(0.0032)	(0.0067)	(0.0037)	(0.0068)	(0.0038)
College educated mother	-0.0158^{*}	-0.0098^{**}	-0.0067	-0.0066	-0.0061	-0.0061
	(0.0082)	(0.0042)	(0.0119)	(0.0055)	(0.0121)	(0.0064)
Primary educated mother	-0.0134^{**}	-0.0098^{***}	-0.0125^{*}	-0.0103^{***}	-0.0108	-0.0098^{***}
	(0.0051)	(0.0027)	(0.0069)	(0.0034)	(0.0067)	(0.0035)
High ranked high school	-0.0153^{**}	-0.0101^{**}	-0.0084	-0.0081	-0.0064	-0.0073
	(0.0068)	(0.0039)	(0.0128)	(0.0061)	(0.0140)	(0.0077)
Low ranked high school	-0.0136^{*}	-0.0084^{**}	-0.0119	-0.0087^{**}	-0.0109	-0.0084^{**}
	(0.0072)	(0.0031)	(0.0081)	(0.0037)	(0.0078)	(0.0040)

This table displays θ coefficients from separate regressions (1). The dependent variable is an indicator equal to one if the student was still in college one year after enrolling. The sample includes students in column (B) of Table 1. The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A) reports estimates of equation (1) using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Column (B) uses the municipality-level treatment variable and municipality dummies. Columns (C)–(D) are analogous to columns (A)–(B), but regressions include dummies for eleven municipality population groups fully interacted with dummies for exam cohort. Columns (E)–(F) include dummies for full interactions between municipality population groups, municipality proximity groups, and exam cohort. See the text and Appendix Table A5 for details on these municipality groups.

Parentheses contain standard errors clustered at the region level.

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

Table A7. Reform effects on college graduation

	(A)	(B)	(C)	(D)	(E)	(F)
	Dependent variable: College graduation					
	Benchmark specification (1)		Within population groups		Within population × proximity groups	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	-0.0153^{*}	-0.0082^{**}	-0.0159	-0.0089^{**}	-0.0146	-0.0087^{*}
	(0.0076)	(0.0039)	(0.0097)	(0.0043)	(0.0099)	(0.0051)
Top income quartile	-0.0271^{***}	-0.0109^{**}	-0.0231^{*}	-0.0099	-0.0223^{*}	-0.0105
	(0.0094)	(0.0051)	(0.0128)	(0.0064)	(0.0130)	(0.0073)
Bottom income quartile	-0.0103	-0.0079^{*}	-0.0161^{*}	-0.0105^{**}	-0.0136	-0.0088^{**}
	(0.0085)	(0.0047)	(0.0086)	(0.0040)	(0.0083)	(0.0039)
College educated mother	-0.0221^{**}	-0.0085	-0.0189	-0.0070	-0.0201	-0.0082
	(0.0100)	(0.0057)	(0.0130)	(0.0067)	(0.0139)	(0.0079)
Primary educated mother	-0.0065	-0.0065	-0.0102	-0.0087^{**}	-0.0087	-0.0071^{*}
	(0.0075)	(0.0040)	(0.0075)	(0.0034)	(0.0078)	(0.0042)
High ranked high school	-0.0193^{**}	-0.0106^{**}	-0.0214	-0.0111	-0.0212	-0.0119
	(0.0074)	(0.0045)	(0.0160)	(0.0080)	(0.0179)	(0.0106)
Low ranked high school	-0.0100	-0.0040	-0.0131	-0.0062	-0.0112	-0.0049
	(0.0106)	(0.0045)	(0.0104)	(0.0039)	(0.0105)	(0.0040)

This table displays θ coefficients from separate regressions (1). The dependent variable is an indicator for college graduation. The sample includes students in column (B) of Table 1. The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A) reports estimates of equation (1) using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Column (B) uses the municipality-level treatment variable and municipality dummies. Columns (C)–(D) are analogous to columns (A)–(B), but regressions include dummies for eleven municipality population groups fully interacted with dummies for exam cohort. Columns (E)–(F) include dummies for full interactions between municipality population groups, municipality proximity groups, and exam cohort. See the text and Appendix Table A5 for details on these municipality groups.

Parentheses contain standard errors clustered at the region level.

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

Table A8. Reform effects on log daily earnings 10–11 years later

	(A)	(B)	(C)	(D)	(E)	(F)
	Dependent variable: Log daily earnings 10–11 years later					
	Benchmark specification (1)		Within population groups		Within population × proximity groups	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	-0.0128^{**}	-0.0075^{**}	-0.0133^{*}	-0.0077^{**}	-0.0175^{***}	-0.0113^{***}
	(0.0053)	(0.0027)	(0.0066)	(0.0036)	(0.0062)	(0.0035)
Top income quartile	-0.0178^{*}	-0.0081	-0.0254^{*}	-0.0100	-0.0341^{***}	-0.0126
	(0.0092)	(0.0055)	(0.0126)	(0.0079)	(0.0119)	(0.0085)
Bottom income quartile	-0.0136^{***}	-0.0079^{**}	-0.0154^{***}	-0.0085^{**}	-0.0168^{**}	-0.0104^{***}
	(0.0047)	(0.0032)	(0.0050)	(0.0036)	(0.0063)	(0.0038)
College educated mother	-0.0111	-0.0038	-0.0140	-0.0005	-0.0276^{*}	-0.0073
	(0.0108)	(0.0059)	(0.0141)	(0.0074)	(0.0138)	(0.0082)
Primary educated mother	-0.0098^{**}	-0.0075^{***}	-0.0112^{***}	-0.0086^{***}	-0.0098^{**}	-0.0090^{**}
	(0.0046)	(0.0025)	(0.0041)	(0.0028)	(0.0044)	(0.0033)
High ranked high school	-0.0060	-0.0062	-0.0064	-0.0055	-0.0126	-0.0071
	(0.0088)	(0.0050)	(0.0132)	(0.0067)	(0.0149)	(0.0094)
Low ranked high school	-0.0125^{**}	-0.0046	-0.0170^{***}	-0.0067^{**}	-0.0184^{***}	-0.0075^{**}
	(0.0055)	(0.0029)	(0.0052)	(0.0029)	(0.0056)	(0.0033)

This table displays θ coefficients from separate regressions (1). The dependent variable is log daily earnings measured 10–11 years after the ICFES exam. The sample includes the subset of students from column (B) of Table 1 who have formal sector earnings. The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A) reports estimates of equation (1) using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Column (B) uses the municipality-level treatment variable and municipality dummies. Columns (C)–(D) are analogous to columns (A)–(B), but regressions include dummies for eleven municipality population groups fully interacted with dummies for exam cohort. Columns (E)–(F) include dummies for full interactions between municipality population groups, municipality proximity groups, and exam cohort. See the text and Appendix Table A5 for details on these municipality groups.

Parentheses contain standard errors clustered at the region level.

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

TABLE A9. Reform effects on log annual earnings in 2017

	(A)	(B)	(C)	(D)	(E)	(F)
	Dependent variable: Log annual earnings in 2017					
	Benchmark specification (1)		Within population groups		Within population × proximity groups	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	-0.0244^{**}	-0.0151^{***}	-0.0242^{*}	-0.0140^{**}	-0.0222	-0.0147^{**}
	(0.0090)	(0.0034)	(0.0137)	(0.0057)	(0.0152)	(0.0066)
Top income quartile	-0.0421^{**}	-0.0211^{**}	-0.0534^{**}	-0.0251^{**}	-0.0514^{*}	-0.0236
	(0.0185)	(0.0087)	(0.0255)	(0.0115)	(0.0273)	(0.0143)
Bottom income quartile	-0.0181	-0.0150	-0.0292^{*}	-0.0201^{*}	-0.0375^{**}	-0.0256^{**}
	(0.0133)	(0.0090)	(0.0172)	(0.0102)	(0.0172)	(0.0109)
College educated mother	-0.0331	-0.0211^{*}	-0.0439	-0.0238^{*}	-0.0526	-0.0285^{*}
	(0.0238)	(0.0111)	(0.0340)	(0.0131)	(0.0383)	(0.0166)
Primary educated mother	-0.0197^{**}	-0.0100^{*}	-0.0178	-0.0076	-0.0198^{**}	-0.0072
	(0.0082)	(0.0053)	(0.0115)	(0.0067)	(0.0091)	(0.0057)
High ranked high school	-0.0241^{*}	-0.0141^{**}	-0.0231	-0.0097	-0.0192	-0.0034
	(0.0130)	(0.0068)	(0.0208)	(0.0102)	(0.0229)	(0.0139)
Low ranked high school	-0.0128	-0.0093	-0.0168	-0.0114	-0.0146	-0.0133^{*}
	(0.0170)	(0.0068)	(0.0185)	(0.0079)	(0.0177)	(0.0069)

This table displays θ coefficients from separate regressions (1). The dependent variable is log annual earnings measured in 2017. The sample includes the subset of students from column (B) of Table 1 who have formal sector earnings in this year. The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A) reports estimates of equation (1) using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Column (B) uses the municipality-level treatment variable and municipality dummies. Columns (C)–(D) are analogous to columns (A)–(B), but regressions include dummies for eleven municipality population groups fully interacted with dummies for exam cohort. Columns (E)–(F) include dummies for full interactions between municipality population groups, municipality proximity groups, and exam cohort. See the text and Appendix Table A5 for details on these municipality groups.

Parentheses contain standard errors clustered at the region level.

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

Table A10. Reform effects on alternative earnings measures (10–11 years later)

	(A)	(B)	(C)	(D)	(E)	(F)
	Dep. variable: Log daily earnings		Dep. variable: Daily earnings (in levels with zeroes)		Dep. variable: Inverse hyperbolic sine of daily earnings	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	-0.013^{**}	-0.007^{**}	-0.600^{***}	-0.371^{***}	-0.047	-0.035^{**}
	(0.005)	(0.003)	(0.189)	(0.084)	(0.035)	(0.013)
Top income quartile	-0.018^{*}	-0.008	-1.279^{***}	-0.691^{***}	-0.091	-0.060^{**}
	(0.009)	(0.006)	(0.381)	(0.175)	(0.061)	(0.022)
Bottom income quartile	-0.014^{***}	-0.008^{**}	-0.294^{**}	-0.230^{***}	-0.024	-0.027^{*}
	(0.005)	(0.003)	(0.132)	(0.072)	(0.028)	(0.015)
College educated mother	-0.011	-0.004	-1.001^{**}	-0.594^{***}	-0.076	-0.057^{**}
	(0.011)	(0.006)	(0.396)	(0.173)	(0.064)	(0.025)
Primary educated mother	-0.010^{**}	-0.007^{***}	-0.294^{**}	-0.250^{***}	-0.014	-0.019
	(0.005)	(0.002)	(0.132)	(0.071)	(0.024)	(0.013)
High ranked high school	-0.006	-0.006	-0.979^{**}	-0.620^{***}	-0.067	-0.051^{*}
	(0.009)	(0.005)	(0.464)	(0.170)	(0.067)	(0.025)
Low ranked high school	-0.013^{**}	-0.005	-0.352^{***}	-0.201^{***}	-0.045^{*}	-0.028^{**}
	(0.005)	(0.003)	(0.121)	(0.061)	(0.025)	(0.012)
N (All students)	340,623	340,623	612,949	612,949	612,949	612,949

Notes: This table displays θ coefficients from separate regressions (1). The dependent variables are daily earnings measured 10–11 years after the ICFES exam, which I define in three different ways. In columns (A)–(B), the dependent variable is log daily earnings, which replicates the benchmark results from column (C) of Table 6. In columns (C)–(D), the dependent variable is daily earnings in levels including zeroes for individuals who do not appear in the formal earnings records. In columns (E)–(F), the dependent variable is the inverse hyperbolic sine of daily earnings (including zeroes).

The sample for columns (C)–(F) includes all college enrollees in the 1998–2001 cohorts (column B of Table 1). The sample for columns (A)–(B) includes only the subset of these students who appear in the formal earnings records. The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A), (C), and (E) report estimates using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Columns (B), (D), and (F) use the municipality-level treatment variable and municipality dummies.

Parentheses contain standard errors clustered at the region level.

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

Table A11. Reform effects on college graduation and earnings in household survey data

	(A)	(B)	(C)	(D)	(E)
	Dep. variable: Graduated from college		Dep. variable: Log hourly earnings		
Definition of birth cohorts exposed to 2000 ICFES reform	All HS grads	College enrollees	All HS grads	College enrollees	No college
1981–1985 birth years	−0.008	-0.014^{*}	−0.011	−0.019	0.003
	(0.009)	(0.008)	(0.010)	(0.013)	(0.015)
1982–1985 birth years	−0.008	-0.017^{**}	−0.013	−0.020	0.000
	(0.009)	(0.007)	(0.010)	(0.014)	(0.015)
1983–1985 birth years	−0.004	-0.012^{*}	−0.012	-0.024^{*}	0.002
	(0.008)	(0.006)	(0.009)	(0.012)	(0.014)
1984–1985 birth years	−0.005	-0.016^{**}	-0.020^{*}	−0.026	−0.013
	(0.008)	(0.007)	(0.011)	(0.016)	(0.015)
N	481,787	247,189	365,259	188,262	176,997

Notes: This table displays estimates of the effects of the 2000 ICFES reform on region-level mean college graduation rates and log hourly earnings using Colombian household survey data. The data are from the January 2015 through June 2019 waves of the household survey *Gran Encuesta Integrada de Hogares*.

In columns (A)–(B), the dependent variable is an indicator for completing a technical college, university, or post-graduate degree. In columns (C)–(E), the dependent variable is log hourly earnings, defined as log(Last month's earnings/(Usual weekly hours × 30/7)).

The sample for columns (A) and (C) includes all surveyed individuals who were born between 1970–1985 and who completed a high school degree. Columns (B) and (D) restrict to the subset of these individuals who enrolled in a technical college or university. Column (E) includes high school graduates who never enrolled in college. Columns (C)–(E) include only individuals with a non-zero hourly wage.

The table displays θ coefficients from the difference-in-differences regression:

$$y_{icrt} = \gamma_c + \gamma_r + \gamma_{a(ct)} + \theta(\text{Treated}_r \times \text{Post}_c) + u_{icrt}$$

where y_{icrt} is an outcome for individual i who was born in year c, lives in region r, and was surveyed in year t. The regression includes birth year fixed effects, γ_c, region fixed effects, γ_r, and age fixed effects, $\gamma_{a(ct)}$, where $a(ct) = t - c$. The coefficient of interest, θ, is on the interaction between a dummy for treated regions, Treated_r, and a variable that denotes birth cohorts that were exposed to the 2000 ICFES exam reform, Post_c. Since most individuals take the ICFES exam between the ages of 16–19, I define $\text{Post}_c = 0$ for $c \leq 1980$ and $\text{Post}_c = 1$ for $c \geq 1984$ in all regressions. I test the sensitivity of the results to defining the 1981–1983 birth years as exposed or unexposed to the reform, as indicated by the row headers. In the first row, I define $\text{Post}_c = 1 - (1984 - c)/4$ for $c \in$ 1981–1983, reflecting the partial exposure of these cohorts. The other rows are similar, except $\text{Post}_c = 0$ for all cohorts outside of the indicated range. All regressions are estimated using survey weights. Parentheses contain standard errors clustered at the region level.

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

Table A12. Clustered and wild t bootstrap p values for reform effects on SES college selectivity gaps

	(A)	(B)
Dependent variable: College selectivity, Q_c		
	Benchmark specification (2)	
	Region-level	Muni-level
Family income gap	−2.402	−1.561
Top Q − Bottom Q	(0.000)	(0.000)
	[0.004]	[0.004]
Mother's education income gap	−1.409	−0.803
College − Primary	(0.007)	(0.015)
	[0.048]	[0.052]
High school rank gap	−1.476	−1.308
High − Low	(0.119)	(0.011)
	[0.180]	[0.020]

Notes: This table presents regression coefficients identical to columns (A)–(B) in Table 5. Parentheses contain p values from standard errors clustered at the region level, as in Table 5. Brackets contain p values from a wild t bootstrap with 500 replications that imposes the null hypothesis, as recommended in Cameron et al. (2008).

TABLE A13. Clustered and wild t bootstrap p values for reform effects on college persistence, graduation, and earnings

	(A)	(B)	(C)	(D)	(E)	(F)
	Dep. variable: One year persistence		Dep. variable: College graduation		Dep. variable: Log daily earnings	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	−0.014	−0.010	−0.015	−0.008	−0.013	−0.007
	(0.009)	(0.002)	(0.051)	(0.043)	(0.021)	(0.010)
	[0.052]	[0.012]	[0.064]	[0.040]	[0.056]	[0.020]
Top income quartile	−0.017	−0.012	−0.027	−0.011	−0.018	−0.008
	(0.065)	(0.007)	(0.007)	(0.042)	(0.063)	(0.150)
	[0.140]	[0.016]	[0.020]	[0.056]	[0.088]	[0.208]
Bottom income quartile	−0.021	−0.013	−0.010	−0.008	−0.013	−0.008
	(0.001)	(0.000)	(0.251)	(0.106)	(0.008)	(0.020)
	[0.020]	[0.036]	[0.307]	[0.160]	[0.024]	[0.072]
College educated mother	−0.016	−0.010	−0.022	−0.008	−0.011	−0.004
	(0.065)	(0.025)	(0.034)	(0.143)	(0.310)	(0.523)
	[0.196]	[0.028]	[0.096]	[0.236]	[0.331]	[0.555]
Primary educated mother	−0.013	−0.010	−0.007	−0.007	−0.010	−0.007
	(0.013)	(0.001)	(0.393)	(0.114)	(0.041)	(0.005)
	[0.080]	[0.024]	[0.371]	[0.136]	[0.068]	[0.004]
High ranked high school	−0.015	−0.010	−0.019	−0.011	−0.006	−0.006
	(0.032)	(0.014)	(0.013)	(0.025)	(0.504)	(0.220)
	[0.112]	[0.044]	[0.028]	[0.028]	[0.563]	[0.287]
Low ranked high school	−0.014	−0.008	−0.010	−0.004	−0.013	−0.005
	(0.068)	(0.011)	(0.354)	(0.387)	(0.029)	(0.114)
	[0.076]	[0.044]	[0.395]	[0.451]	[0.032]	[0.112]

Notes: This table presents regression coefficients identical to those in Table 6. Parentheses contain p values from standard errors clustered at the region level, as in Table 6. Brackets contain p values from a wild t bootstrap with 500 replications that imposes the null hypothesis, as recommended in Cameron et al. (2008).

Table A14. Reform effects on SES college selectivity gaps excluding certain regions
Dependent variable: College selectivity, Q_c

	(A)	(B)	(C)
	All regions	Exclude small regions	Exclude mixed admit
Family income gap	-2.40^{***}	-2.47^{***}	-2.01^{**}
Top Q − Bottom Q	(0.54)	(0.54)	(0.74)
Mother's education gap	-1.41^{***}	-1.44^{***}	-0.90
College − Primary	(0.49)	(0.50)	(0.65)
High school rank gap	-1.48	-1.54	-0.65
High − Low	(0.92)	(0.94)	(1.07)

Notes: This table displays θ^q coefficients from separate regressions (2). The sample includes students in Column (B) of Table 1. Rows are defined by the three SES measures from Table 1.

Columns (A) reports estimates from equation (2) using the region-level treatment variable (Treatment_g) and fixed effects for regions (γ_g and γ_{gt}). Column (B) is analogous to column (A), but it omits the seven regions with no public universities (Casanare, Guainía, Guaviare, Putumayo, San Andrés, Vaupés, and Vichada). Column (C) is analogous to column (B), but it also omits the three regions with mixed admission methods (Bogotá, Caldas, and Valle del Cauca). See Appendix Table A3 for details on the regions.

Parentheses contain standard errors clustered at the region level.

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

TABLE A15. Reform effects on college graduation and earnings excluding certain regions

	(A)	(B)	(C)	(D)	(E)	(F)
	Dep. variable: College graduation			Dep. variable: Log daily earnings		
	All regions	Exclude small regions	Exclude mixed admit	All regions	Exclude small regions	Exclude mixed admit
All students	-0.015^{*}	-0.015^{*}	-0.015	-0.013^{**}	-0.014^{**}	-0.016^{*}
	(0.008)	(0.008)	(0.013)	(0.005)	(0.005)	(0.009)
Top income quartile	-0.027^{***}	-0.027^{***}	-0.028^{*}	-0.018^{*}	-0.020^{**}	-0.023
	(0.009)	(0.009)	(0.015)	(0.009)	(0.009)	(0.016)
Bottom income quartile	-0.010	-0.010	-0.012	-0.013^{***}	-0.015^{***}	-0.015^{**}
	(0.009)	(0.009)	(0.013)	(0.005)	(0.004)	(0.006)
College educated mother	-0.022^{**}	-0.022^{**}	-0.019	-0.011	-0.013	-0.012
	(0.010)	(0.010)	(0.017)	(0.011)	(0.011)	(0.020)
Primary educated mother	-0.007	-0.006	-0.004	-0.010^{**}	-0.012^{***}	-0.013^{**}
	(0.008)	(0.008)	(0.011)	(0.005)	(0.004)	(0.005)
High ranked high school	-0.019^{**}	-0.019^{**}	-0.023	-0.006	-0.008	-0.011
	(0.007)	(0.007)	(0.017)	(0.009)	(0.009)	(0.015)
Low ranked high school	-0.010	-0.009	-0.012	-0.013^{**}	-0.014^{**}	-0.017^{**}
	(0.011)	(0.011)	(0.013)	(0.005)	(0.005)	(0.006)

Notes: This table displays θ coefficients from separate regressions (1) with two dependent variables: an indicator for college graduation (columns A–C); and log daily earnings 10–11 years after the admission exam (columns D–F). The sample in columns (A)–(C) includes students in Column (B) of Table 1; columns (D)–(F) include only students with formal sector earnings. The first row shows estimates for all students; other rows report estimates by SES group.

All columns report estimates using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Columns (A) and (D) include all regions. Columns (B) and (E) omit the seven regions with no public universities (Casanare, Guainía, Guaviare, Putumayo, San Andrés, Vaupés, and Vichada). Columns (C) and (F) omit these seven regions plus the three regions with mixed admission methods (Bogotá, Caldas, and Valle del Cauca). See Appendix Table A3 for details on the regions.

Parentheses contain standard errors clustered at the region level.

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

Table A16. Longer-run effects of 2000 ICFES reform

	(A)	(B)	(C)	(D)	(E)	(F)
	Reform effects on correlations with HS mean family income			Reform effects on mean outcomes for all college enrollees		
Exam cohorts	ICFES math score	ICFES physics score	College selectivity	Persisted for one year	College grad	Log annual earnings in 2017
2000–2001 cohorts	-0.174^{***}	-0.184^{***}	-0.017^{**}	-0.013^{**}	-0.014^{*}	-0.019^{**}
	(0.007)	(0.010)	(0.007)	(0.005)	(0.007)	(0.008)
2002–2003 cohorts	-0.131^{***}	-0.149^{***}	-0.021^{*}	-0.013	-0.014	-0.026^{*}
	(0.005)	(0.006)	(0.011)	(0.009)	(0.010)	(0.014)
2004–2005 cohorts	-0.135^{***}	-0.173^{***}	-0.028^{*}	-0.003	-0.017	-0.014
	(0.005)	(0.009)	(0.014)	(0.014)	(0.014)	(0.023)
2006–2007 cohorts	-0.098^{***}	-0.168^{***}	-0.041^{**}	-0.013	-0.020	-0.013
	(0.007)	(0.008)	(0.016)	(0.015)	(0.019)	(0.028)
N	3,487,050	3,487,050	1,597,319	1,597,319	1,597,319	917,124

Notes: This table shows the longer-run effects of the 2000 ICFES reform up to the 2007 exam cohort.

Columns (A)–(B) show how the reform affect correlations of SES with test scores. These regressions are similar to those in Table 2, but I use mean family income at the high school level as my measure of SES since I do not observe family income in all cohorts. I first compute family income in the pre-reform cohorts (1998–1999) using the same method as described in Section 2.2. I then compute the average family income in each high school and use this average for all cohorts from 1998–2007. I regress math (column A) and physics (column B) scores on cohort dummies, HS mean family income, and the interaction of HS mean family income with dummies for the four cohort groups listed in the first column (2000–2001, 2002–2003, 2004–2005, and 2006–2007). Columns (A)–(B) display the coefficients on the interaction variables. The sample includes all ICFES exam takers in 1998–2007.

Column (C) shows how the reform affected the correlation between SES and college selectivity. This regression is similar to those in Table 5, but I again use HS mean family income as my measure of SES. I estimate equation (2) using the region-level treatment variable (Treated_g), but I replace X_i with HS mean family income, and I replace Post_t with dummies for the four cohort groups listed in the first column. Column (C) displays the θ^q coefficients on these triple interactions. The sample includes all college enrollees in 1998–2007.

Columns (D)–(F) show how the reform affected college persistence, college graduation, and earnings for all college enrollees. These regressions are similar to those in the first row of Table 6. I estimate equation (1) using the region-level treatment variable (Treated_g), but I replace Post_t with dummies for the four cohort groups listed in the first column. Columns (D)–(F) display the θ coefficients on these interactions using three different dependent variables: an indicator equal to one if the student was still in college one year after enrolling (column D); an indicator for college graduation by 2012 (column E); and log annual earnings in 2017 (column F). The sample includes all college enrollees in 1998–2007 and the subset of those with formal earnings.

In all columns, parentheses contain standard errors clustered at the region level.

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

TABLE A17. Reform effects on field of study choice

	(A)	(B)	(C)	(D)	(E)	(F)
	Dep. variable: Enrolled in a STEM program		Dep. variable: Enrolled in a Business program		Dep. variable: Enrolled in a health or humanities prog.	
	Region-level	Muni-level	Region-level	Muni-level	Region-level	Muni-level
All students	0.002	0.003	0.003	−0.000	−0.006	−0.002*
	(0.006)	(0.002)	(0.006)	(0.003)	(0.005)	(0.001)
Top income quartile	−0.013	−0.006	0.009	0.006	0.003	0.000
	(0.013)	(0.006)	(0.011)	(0.005)	(0.007)	(0.002)
Bottom income quartile	0.006	0.009***	0.002	−0.004	−0.008	−0.005**
	(0.007)	(0.003)	(0.007)	(0.003)	(0.006)	(0.002)
College educated mother	−0.011	−0.005	0.007	0.004	0.004	0.001
	(0.012)	(0.005)	(0.010)	(0.004)	(0.009)	(0.003)
Primary educated mother	0.005	0.007***	0.003	−0.003	−0.008	−0.003
	(0.006)	(0.002)	(0.006)	(0.003)	(0.006)	(0.002)
High ranked high school	−0.006	−0.003	0.004	0.003	0.002	0.000
	(0.011)	(0.005)	(0.010)	(0.005)	(0.006)	(0.002)
Low ranked high school	−0.001	0.004	0.007	−0.002	−0.006	−0.003
	(0.007)	(0.003)	(0.007)	(0.003)	(0.007)	(0.003)

Notes: This table displays θ coefficients from separate regressions (1). The dependent variables are indicators for enrolling in the field of study group listed in the column header, which are defined by the Ministry of Education's eight program areas. STEM includes the engineering and natural science program areas. Health & humanities includes the health, fine arts, education, social sciences, and agronomy program areas. The mean pre-reform enrollment rate in each field group is: STEM (34.9 percent); business (27.1 percent); and health & humanities (38.0 percent).

The sample includes students in Column (B) of Table 1. The first row shows estimates for all students; other rows report estimates by SES group.

Columns (A), (C), and (E) report estimates using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Columns (B), (D), and (F) use the municipality-level treatment variable and municipality dummies.

Parentheses contain standard errors clustered at the region level.

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

B. Theoretical appendix

This appendix presents a framework that illustrates how the redesign of a college admission exam can impact earnings outcomes. It also uses this framework to discuss how the effects of such reforms can differ from estimates of the returns to college selectivity in related literature.

B.1. **Admission exams and college choice.** I consider a set of college applicants indexed by $i = 1, \ldots, I$. Applicants are characterized their ability, A_i, and by their socioeconomic status (SES), X_i. To apply to college, applicants take a national admission exam. I denote the potential admission test score of individual i who takes the exam in cohort c by

$$T_{ic} = t_c\Big(A_i, X_i, \nu_{ic}\Big). \tag{B1}$$

Test scores depend on ability, SES, and a noise term, ν_{ic}, that captures factors like guessing. Importantly, scores also depend on the design of the exam in an individual's cohort, which affects the mapping from characteristics to scores, $t_c(\cdot)$. Thus T_{ic} is a *potential* outcome because the same student can earn different scores in different cohorts. I consider T_{ic} to be in percentiles so that the unconditional distribution of scores is the same in each cohort, and because an individual's rank relative to other applicants is what matters for admissions.

Applicants can apply to a large number of college programs indexed by $j = 1, \ldots, J$. Individuals' preferences for college programs depend on their potential earnings outcomes and on other non-pecuniary factors. The potential earnings of individual i from enrolling in program j are given by

$$Y_{ic}(j) = y_j\Big(A_i, X_i, I_{jc}, \epsilon_{ij}\Big). \tag{B2}$$

Earnings depend on ability, SES, and an idiosyncratic error, ϵ_{ij}. I allow for peer effects (broadly defined) by letting earnings depend on the set of other students who attend program j in cohort c, which I denote by I_{jc}. I let ω_{ij} denote all other factors that affect an individual's preferences for program j, including tuition, location, and enjoyment of the material.

Individuals enroll in the college program that maximizes their expected utility among those that they can gain admission to. I let $S_{ic} \in \{1, \ldots, J\}$ denote the program that individual i would attend if they took the exam in cohort c. This choice is given by:

$$S_{ic} = \underset{j}{\operatorname{argmax}}\, EU\Big[Y_{ic}(j), \omega_{ij}\Big] \quad \text{s.t.} \quad T_{ic} \geq T^*_{jc}, \tag{B3}$$

where T^*_{jc} is the cutoff score for admission to program j in cohort c. A program's cutoff score can differ across exam cohorts because it depends on distribution of scores in each cohort.

B.2. **Earnings effects of exam reform.** I consider a redesign of the admission exam that reduces the test score gap between high and low SES students. For simplicity, assume individuals are either high SES ($X_i = H$) or low SES ($X_i = L$), and that there are only two cohorts: pre-reform ($c = 0$) and post-reform ($c = 1$). Low SES students score better on the post-reform exam than on the pre-reform exam, i.e.,

$$E[T_{i1}|X_i = L] > E[T_{i0}|X_i = L]. \quad \text{(B4)}$$

Equivalently, the reform reduces the average (percentile) score for high SES students.

A key question is how such reforms also affect the distribution of scores with respect to ability, A_i. Exam designers often aim to reduce the importance of "test prep," which may reduce SES score gaps while also improving the exam's measurement of abilities that are important to succeed in college. But SES gaps can also decline because noise, ν_{ic}, becomes relatively more important. For example, if the exam becomes too hard or too easy, much of the variation in scores is driven by guessing. In this case the exam's predictive power for ability may decrease. The relationship between test scores and ability matters if ability is an important source of heterogeneity in the returns to college selectivity, as I discuss below.

This reform will tend to expand access to college programs for low SES students, and restrict access for high SES students. Let $\succ_i$ denote individual i's preferences over college programs as defined in equation (B3). For low SES students, the fraction of individuals who would gain access to a preferred program if they took the new exam ($S_{i1} \succ_i S_{i0}$) is likely to be larger than the fraction who would lose access to a preferred program ($S_{i1} \prec_i S_{i0}$), i.e.,

$$Pr\Big[S_{i1} \succ_i S_{i0}\Big|X_i = L\Big] > Pr\Big[S_{i1} \prec_i S_{i0}\Big|X_i = L\Big]. \quad \text{(B5)}$$

Conversely, high SES students will more often lose access to a preferred program than gain access. But not all low SES students benefit because the exam reform can also affect the importance of ability or noise. Some low SES students may lose access to preferred programs; others may attend the same program regardless of their exam cohort.

My empirical analysis below estimates the earnings effects of exam reforms that reduce SES test score gaps. The mean earnings effect across all students is given by $E[Y_i|C_i = 1] - E[Y_i|C_i = 0]$, where Y_i is an individual's realized earnings and $C_i \in \{0, 1\}$ is their realized exam cohort.[32] I also consider effects on earnings inequality by computing the reform effect conditional on X_i.

To illustrate the mechanisms through which the exam reform can affect earnings, I start with the simple case that assumes no peer effects. If earnings do not depend on college peer composition, I_{jc}, then potential earnings can be written as $Y_i(j)$ without the cohort

[32] Identification of the reform effect relies on the assumption that C_i is unrelated to an individual's potential outcomes for both program choice, S_{ic}, and earnings, $Y_{ic}(j)$. In my empirical analysis I make a difference-in-differences version this assumption that requires parallel trends in these outcomes.

subscript, c. In this case, the mean earnings effect of the reform can be written as:

$$
\begin{aligned}
E[Y_i|C_i=1]-E[Y_i|C_i=0]= \quad & E\Big[Y_i(S_{i1})-Y_i(S_{i0})\Big|S_{i1}\succ_i S_{i0}\Big]Pr\Big[S_{i1}\succ_i S_{i0}\Big] \\
& -E\Big[Y_i(S_{i0})-Y_i(S_{i1})\Big|S_{i1}\prec_i S_{i0}\Big]Pr\Big[S_{i1}\prec_i S_{i0}\Big].
\end{aligned} \tag{B6}
$$

Without peer effects, the change in earnings is driven only by students who attend different college programs because of the reform. Equation (B6) shows that this effect depends on the relative returns to attending a preferred college program for two groups of students. Mean earnings increase if the students who are "shifted in" to a preferred program ($S_{i1} \succ_i S_{i0}$) have larger returns to those programs than the students who are "shifted out" of their preferred programs ($S_{i1} \prec_i S_{i0}$), weighted by the proportion of students in each group. Mean earnings decrease if the students in the "shifted out" group have larger returns in aggregate.

The analysis is more complicated if earnings depend on peer composition, I_{jc}, because the exam reform can also affect earnings for students who attend the same program in either cohort ($S_{i1} = S_{i0}$). In this case, the mean earnings effect still depends on the relative returns for "shifted in" and "shifted out" students, but it also depends on any spillover effects of peer composition on the returns of any student, i.e., $y_j(I_{j0}, \dots) \neq y_j(I_{j1}, \dots)$. For example, peer effects could arise if students benefit from attending colleges in which their own ability or SES is similar to that of their classmates. There is empirical evidence that spillover effects may be important for large-scale admission reforms, as I discuss below.

B.3. **Heterogeneity by SES and ability.** There are many reasons why students who gain and lose access to preferred programs from an exam reform could differ in their returns. Research on the returns to college selectivity often finds significant heterogeneity with respect to SES, X_i. Some work finds that low-SES or minority students have *larger* returns to selective colleges than students from more advantaged backgrounds (Dale and Krueger, 2002; Bleemer, 2022), possibly because they benefit more from these schools' financial resources. Other papers find that high-SES students have larger returns to top colleges because they benefit more from peer networks (Zimmerman, 2019; Michelman et al., 2022). Thus reforms that reduce SES test score gaps could either raise or lower mean earnings, depending on which of these mechanisms dominates for the set of students and colleges that are impacted.

Other research emphasizes the importance of heterogeneity with respect to ability, A_i. Some work finds that students have higher graduation rates and earnings when their own academic preparation is close to that of their classmates (Arcidiacono et al., 2016; Dillon and Smith, 2020), suggesting that ability and college selectivity are complementary. A special case of complementarity is the "mismatch hypothesis" (Arcidiacono and Lovenheim, 2016), which states that students with less academic preparation might have *lower* earnings if they attend a more selective school. In my framework, mismatch implies that low-SES students have *negative* returns to attending a preferred program, i.e., $E\Big[Y_i(S_{i1})-Y_i(S_{i0})\Big|X_i = L, S_{i1} \succ_i S_{i0}\Big] < 0$. This can arise if students have strong preferences for non-pecuniary aspects of programs, ω_{ij}, that outweigh the negative earnings returns, or if they have incorrect expectations on their returns (Arcidiacono et al., 2011).

If there is heterogeneity by ability, it is important to understand how exam redesigns also affect the measurement of ability. Reforms that reduce the exam's predictive power for ability can reduce mean earnings if ability and college selectivity are complementary. If mismatch is possible, earnings can fall for *both* high- and low-SES students.

Importantly, my framework shows that the "mismatch hypothesis" is not the only reason why an exam reform could reduce the earnings of low-SES students. Even within the population of low-SES students, the earnings effect of the reform still depends on the relative returns for "shifted in" and "shifted out" students, as in equation (B6).[33] Although a greater fraction of low-SES students are in the "shifted in" group than in the "shifted out" group (equation B5), the reform effect can be negative if the "shifted out" group has particularly large returns to their preferred programs. This is a different type of mismatch; even if all students have positive returns, mismatch can arise if the low-SES students who would particularly benefit from attending certain programs do not gain admission to those programs.

B.4. **Empirical strategies in the literature.** Given the many potential sources of heterogeneity, it is important to have empirical evidence on how exam reforms affect earnings outcomes. Although there is a large literature on the returns to college selectivity, these papers may not be that informative for the effects of an admission exam reform. By the nature of their identification strategies, most papers typically estimate returns for students who are *qualified* for admission to a selective college. Many papers use regression discontinuity designs based on admission thresholds (Hoekstra, 2009; Kirkebøen et al., 2016), and thus estimated returns come from students on the margin of admission. Other papers use identification strategies that compare students who were admitted to the same colleges but made different enrollment choices (Dale and Krueger, 2002; Mountjoy and Hickman, 2020). By contrast, equation (B6) shows that the earnings effect of an exam reform depends on precisely those students who would *not* be qualified for admission under the old exam.

Further, estimates from the literature typically do not include spillover effects, which may be important in large-scale admission reforms. In most papers, college peer composition is either implicitly or explicitly held fixed in the authors' identification strategy (Chetty et al., 2020). But there is evidence that peer composition affects student outcomes through

[33] Specifically, the earnings effect for low SES students, $E[Y_i|X_i = L, C_i = 1] - E[Y_i|X_i = L, C_i = 0]$, can be decomposed in the same way as in equation (B6) by conditioning on $X_i = L$ in all righthand side terms.

channels of learning spillovers (Arcidiacono and Vigdor, 2010) or peer and alumni networking (Machado et al., 2022).

The paper that is most similar to my setting is Black et al. (2023), which examines earnings outcomes for students who were "pulled in" to and "pushed out" of selective universities by Texas' Top 10 Percent policy. The "pulled in" group included students from high schools that were historically underrepresented at top colleges, while the "pushed out" group consisted of students from traditional feeder schools. The authors show that the policy led to increases in degree attainment and earnings for the "pulled in" group, and that it did not have significant effects on outcomes for the "pushed out" group. Crucially, the authors find that the "pulled in" group had *higher* ability than the "pushed out" group as measured by high school test scores. As I show below, students who were "pulled in" to selective Colombian universities by the exam reform had *lower* ability as measured by the exams' predictive power for college success. Thus our settings differ in the characteristics of the $S_{i1} \succ_i S_{i0}$ and $S_{i1} \prec_i S_{i0}$ groups, which can lead the admission reforms to have different impacts. This motivates my empirical analysis of how the Colombian admission exam reform impacted the distributions of test scores, college selectivity, and earnings.

C. Empirical appendix

Table C1. Construction of analysis sample

	N
Total number of exam takers	2,100,424
Missing exam scores	(10,195)
Missing high school information	(126,966)
Fewer than five pre-reform obs. in gender/HS/mom ed cells	(319,003)
Full sample	1,644,260
College enrollees	612,949

Notes: See the text for descriptions of the sample restrictions in each row.

C.1. **Data, sample, and variable definitions.** This section describes the coverage and merging of my three main administrative datasets: college admission exam records, enrollment and graduation records, and earnings records. It also describes my analysis sample and definitions of key variables.

The first dataset includes records from the ICFES national standardized college entrance exam. The data include all students who took the exam between 1998–2001. My sample includes all exam takers with non-missing test scores and high school identifiers. In addition, I make a sample restriction that allows me to calculate a consistent measure of family income across cohorts. In the data, family income is grouped into ten bins based on multiples of the monthly minimum wage, but the distribution of these bins changes dramatically across cohorts due to variation in inflation and minimum wage policy. To get a stable measure of family income, I calculate predicted family income using an individual's gender, mother's education, and high school. Specifically, I define predicted income as the mean family income as fraction of the minimum wage within cells defined by gender, nine mother's education categories, and the roughly 7,500 high schools in my sample. I use only 1998–1999 cohorts to predict family income, and I drop any cells with fewer than five observations in the pre-reform cohorts. I then define income quartiles based on a student's percentile rank of predicted family income within their exam cohort.

Table C1 shows the effect of these restrictions on sample size. The full sample includes roughly 1.6 million exam takers. Table C2 shows that I find no evidence of differential selection into the sample in treatment and control areas. Most of my analyses are restricted to those who enrolled in college, as shown in the last row of Table C1. I find no evidence of reform effects on the probability of college enrollment (Appendix Table A4).

In addition to family income quartiles, I use two other SES measures computed from the ICFES records. First, I use two mother education groups: students whose mothers

Table C2. Balance test for missing data
Dependent variable: Appears in analysis sample

	(A) Region-level	(B) Muni-level
All students	−0.006	−0.001
	(0.008)	(0.004)
N	1,962,923	1,962,923
Mean	0.838	0.838

Notes: This table displays θ coefficients from separate regressions (1). The dependent variable is an indicator equal to one if the student appears in the analysis sample (fifth row of Table C1). The sample includes all exam takers except those with missing exam scores and missing high school information (rows 2–3 of Table C1).

Column (A) reports estimates using the region-level treatment variable, Treatment_g, and region dummies, γ_g. Column (B) uses the municipality-level treatment variable and municipality dummies.

Parentheses contain standard errors clustered at the region level.

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

attended college (or more) and students whose mothers attended primary school or less. Second, I use groups defined by high school rank. The ICFES testing agency classifies high schools into seven categories based on their mean admission exam performance. The low rank group includes the bottom three categories, and the high rank group includes the top three categories. I use a high school's pre-reform rank and hold this definition fixed across cohorts.

The second dataset includes enrollment and graduation records from the Ministry of Education. The Ministry's records include almost all colleges in Colombia, although it omits a few schools due to their small size or inconsistent reporting. To describe the set of colleges that are included in the Ministry of Education records, I use another administrative dataset from a college exit exam called *Saber Pro* (formerly ECAES). This national exam is administered by the same agency that runs the ICFES college admission exam. The exit exam became a requirement for graduation from any higher education institution in 2009.

Column (A) in Table C3 depicts the 310 colleges that have any exit exam takers in these administrative records in 2009–2011. These colleges are categorized into the Ministry of Education's five types of higher education institutions, which are listed in descending order of their on-time program duration.[34] Column (B) shows the number of exit exam takers per year. The majority of exam takers are from university-level institutions, with fewer students from technical colleges.

Column (C) shows the fraction of these 310 colleges that appear in the Ministry of Education records that I use in my analysis. These proportions are weighted by the number

[34] Most programs at universities require 4–5 years of study, while programs at Technical/Professional Institutes typically take 2–3 years.

Table C3. Higher education institutions in Ministry of Education records

	(A) Number of colleges	(B) Number of exit exam takers/year	(C) Prop. of colleges in records
University	122	134,496	1.00
University Institute	103	53,338	0.88
Technology School	3	2,041	1.00
Technology Institute	47	15,092	0.82
Technical/Professional Institute	35	11,408	0.99
Total	310	216,375	0.96

Notes: Column (A) depicts the number of colleges that have *Saber Pro* exit exam takers in 2009–2011 using administrative records from the testing agency. Colleges are categorized into the Ministry of Education's five higher education institution types. Column (B) shows the number of 2009–2011 exam takers per year. Column (C) shows the proportion of colleges that appear in the Ministry of Education records, where colleges are weighted by the number of exit exam takers.

of exam takers depicted in column (B). Column (C) shows that the Ministry of Education records include all universities but are missing a few technical colleges.[35] Overall, 96 percent of exit exam takers attend colleges that appear in the Ministry of Education records.

I define my measure of college selectivity, Q_c, as a college's percentile rank in each exam cohort based on the mean pre-reform exam score in each college c.[36] For the 20 colleges with fewer than ten pre-reform enrollees, I define Q_c as the mean exam percentile of the students at all 20 of these colleges.

My last data source is from the Ministry of Social Protection. These data provide monthly earnings for any college enrollee employed in the formal sector in 2008–2012. From these records I calculate the log of an individual's average daily earnings measured 10–11 years after taking the admission exam. Ten and eleven years after the test are the two experience levels at which I can measure earnings for each of the 1998–2001 exam cohorts. I compute average daily earnings by dividing total annual earnings by the number of formal employment days in the year, demeaning by exam cohort and year, and averaging across the two years.

I merge these three datasets using national ID numbers, birth dates, and names. Nearly all students in these records have national ID numbers, but Colombians change ID numbers around age 17. Most students in the admission exam records have the below-17 ID number (*tarjeta*), while the majority of students in the college enrollment and earnings records have the above-17 ID number (*cédula*). Merging using ID numbers alone would therefore lose a large majority of students. Instead, I merge observations with either: 1) the same ID number

[35] The largest omitted institutions are the national police academy (*Dirección Nacional de Escuelas*) and the Ministry of Labor's national training service (*Servicio Nacional de Aprendizaje*).
[36] This measure is the average score across all exam subjects.

and a fuzzy name match; 2) the same birth date and a fuzzy name match; or 3) an exact name match for a name that is unique in both records.

38 percent of the 1998–2001 exam takers appear in the enrollment records, which is comparable to the higher education enrollment rate in Colombia during the same time period.[37] A better indicator of merge success is the percentage of college enrollees that appear in the admission exam records because all domestic college students must take the exam. I match 88 percent of enrollees who took the admission exam between 1998 and 2001.[38]

The exam validity results in Table 3 use college GPA computed from transcript records from one public flagship university (name withheld). I obtained transcript records for students in business, engineering, and architecture programs in the 2000–2004 enrollment cohorts. I calculate first-year GPA as the average grade across all courses the student took in the first two semesters after enrolling.

C.2. **The 2000 ICFES exam reform.** This section provides further details on the 2000 reform of the ICFES admission exam.

The goal of the 2000 ICFES overhaul was to design an exam that supported the dual goals of measuring high school quality and aiding in college admissions. The pre-reform exam was thought to primarily test intellectual ability and rote memorization, and was thus poorly suited for measuring the contribution of high schools to students' educational development. Furthermore, the exam was criticized for being biased toward certain students depending on their gender or family background.

To achieve this goal, the testing agency rewrote the exam with the aim of testing "competencies" rather than "content." The focus of the new was to test "know-how in context," which means that students should be able to apply a given piece of information to different situations. Examples of such competencies include interpreting a text, graphic, or map in solving a problem, and assessing different concepts and theories that support a decision. The post-reform exam therefore placed a greater emphasis on communication skills, as it asked students to interpret, argue, and defend their answers.

[37] The gross tertiary enrollment rate ranged from 22 percent to 24 percent between 1998 and 2001 (World Bank World Development Indicators, available at http://data.worldbank.org/country/colombia in October 2016). This rate is not directly comparable to my merge rate because not all high school aged Colombians take the ICFES exam. About 70 percent of the secondary school aged population was enrolled in high school in this period. Dividing the tertiary enrollment ratio by the secondary enrollment ratio gives a number roughly comparable to my 38 percent merge rate.

[38] The enrollment records contain age at time of the admission exam for some students, which allows me to calculate the year they took the exam. Approximately 16 percent of students in the enrollment dataset have missing birth dates, which accounts for the majority of observations I cannot merge. Some duplicate matches arise because students took the admission exam more than once, though I erroneously match a small number of students with the same birth date and similar names.

Figures C1–C4 present sample questions from the biology, language, math, and social sciences components of the pre-reform and post-reform exams. These sample questions were distributed by the testing agency to describe the central motivation of the overhaul. Questions from the pre-reform exam are briefer and require more memorization. The post-reform sample questions are longer and often include a figure or passage that the student must interpret. Further, some pre-reform questions have a complicated answer structure, while the post-reform questions are all straightforward multiple choice.

These communication and interpretation skills were tested in the context of subjects from the core secondary education curriculum. To better align the test with the high school curriculum, the reform also altered the specific subjects that were tested. Table A2 shows the subject components that were included in the admission exam between 1998 and 2001. The 2000 reform combined two math exams—one designed to measure aptitude and another designed to test knowledge—into a single component. The reform also split the social sciences component into separate tests for history and geography. Further, the 2000 reform added components in philosophy and foreign language, which was English for the large majority of students.

In Section 2 I focus on the six subject groups listed in the leftmost column of Table A2: biology, chemistry, language, math, physics, and social sciences. I average the pre-reform math aptitude and math knowledge components into a single math score. I also average the post-reform history and geography components into a single social sciences score. I exclude the verbal component, which appears only in the pre-reform exam, and the philosophy and foreign language components, which appear only in the post-reform exam. I also exclude the elective component, which was rarely used by colleges to determine admissions.

Table A2 also shows that the reform affected score means, standard deviations, and number of unique values. The bottom rows show that the mean score across all subjects was about 50 in the pre-reform cohorts, and 45 in the post-reform cohorts. The standard deviation across all components fell from approximately ten to six in most subjects. Lastly, the reform also reduced the number of unique score values per exam administration from roughly 50 to 40.

Panel A. Pre-reform sample question

Which of the following two are inverse chemical processes?

1. Photosynthesis
2. Cyclosis
3. Breathing
4. Circulation

(A) If 1 and 2 are correct, fill in oval A
(B) If 2 and 3 are correct, fill in oval B
(C) If 3 and 4 are correct, fill in oval C
(D) If 2 and 4 are correct, fill in oval D
(E) *If 1 and 3 are correct, fill in oval E*

Panel B. Post-reform sample question

The diagram shows a cell that is exchanging substances with its environment through the cell membrane.

If at a certain time it is observed that the number of molecules A entering the cell is greater than the number coming out of it, it can be assumed that within the cell there is

(A) A higher concentration of molecules than outside
(B) *A lower concentration of molecules than outside*
(C) A molecule concentration equal to that outside
(D) An absence of molecules A

FIGURE C1. Biology sample questions

Notes: Correct answers are in *italics*. The sample question from the pre-reform exam was obtained from a version of the ICFES testing agency's website that was archived in January 1997 (available at https://web.archive.org/web/19980418191357/http://acuario.icfes.gov.co/12/122/1222/12223/Tipos.html in October 2016). The sample question from the post-reform exam was obtained from a September 2008 ICFES report entitled "State Assessment Tests in Colombia" (*"Evaluación con Pruebas de Estado en Colombia"*) (available at http://www.ieia.com.mx/materialesreuniones/1aReunionInternacionaldeEvaluacion/PONENCIAS18Septiembre/ConferenciasMagnas/MargaritaPenaBorrero.pdf in October 2016).

Panel A. Pre-reform sample question

The phrase:

"¿Estará Pedro en la casa?"

is used to ask about the location of Pedro:

(A) *At the moment when the question is asked*
(B) At a future moment
(C) At any moment
(D) At the moment when the answer is given

Panel B. Post-reform sample question

Me parece que no es preciso demostrar que la novela policial es popular, porque esa popularidad es tan flagrante que no requiere demostración. Para explicarla—aquellos que niegan al género su significación artística—se fundan en la evidencia de que la novela policial ha sido y es uno de los productos predilectos de la llamada "cultura de masas," propia de la moderna sociedad capitalista.

La popularidad de la novela policial sería, entonces, sólo un resultado de la manipulación del gusto, sólo el fruto de su homogeneización mediante la reiteración de esquemas seudoartísticos, fácilmente asimilables, y desprovistos, claro, de verdadera significación gnoseológica y estética; sazonados, además, con un puñado de ingredientes de mala ley: violencia, morbo, pornografía, etcétera, productos que se cargan, casi siempre, de mistificaciones y perversiones ideológicas, tendientes a la afirmación del estatus burgués y a combatir las ideas revolucionarias y progresistas del modo más burdo e impúdico.

Pero hay que decir que ello constituye no sólo una manipulación del gusto en general, sino también una manipulación de la propia novela policial, de sus válidas y legítimas manifestaciones, una prostitución de sus mecanismos expresivos y sus temas. Los auténticos conformadores del género policial (no hay que olvidarlo) fueron artistas de la talla de Edgar Allan Poe y Wilkie Collins. Y desde sus orígenes hasta nuestros días, el género ha producido una buena porción de obras maestras.

From "La novela policial y la polémica del elitismo y comercialismo"
In *Ensayos Voluntarios*, Guillermo Rodríguez Rivera.
Havana, *Editorial Letras Cubanas*, 1984.

The theme of the previous text is:

(A) The pseudo-artistic nature of detective novels is devoid of epistemological and aesthetic significance
(B) The detective novel is a favorite product of the so-called "mass culture"
(C) The popularity of the detective genre is not necessary to show through evidence
(D) *Detective novels and their manifestations can manipulate tastes*

FIGURE C2. Language sample questions

Notes: Correct answers are in *italics*. The sample question from the pre-reform exam was obtained from a version of the ICFES testing agency's website that was archived in January 1997 (available at https://web.archive.org/web/19980418191357/http://acuario.icfes.gov.co/12/122/1222/12223/Tipos.html in October 2016). The sample question from the post-reform exam was obtained from a September 2008 ICFES report entitled "State Assessment Tests in Colombia" (*"Evaluación con Pruebas de Estado en Colombia"*) (available at http://www.ieia.com.mx/materialesreuniones/1aReunionInternacionaldeEvaluacion/PONENCIAS18Septiembre/ConferenciasMagnas/MargaritaPenaBorrero.pdf in October 2016).

Panel A. Pre-reform sample question

It is known that the result of multiplying a number by itself several times is 256. You can identify this number if it is known

I. Whether the number is positive or negative
II. How many times the number is multiplied by itself

(A) If fact I is enough to solve the problem, but fact II is not, fill in oval A
(B) If fact II is enough to solve the problem, but fact I is not, fill in oval B
(C) *If facts I and II together are sufficient to solve the problem, but each separately it is not, fill in oval C*
(D) If each of facts I and II separately are sufficient to solve the problem, fill in oval D
(E) If facts I and II together are not enough to solve the problem, fill in oval E

Panel B. Post-reform sample question

To test the effect of a vaccine applied to 516 healthy mice, an experiment was performed in a laboratory. The goal of the experiment is to identify the percentage of mice that become sick when subsequently exposed to a virus that attacks the vaccine. The following graphs represent the percentage of sick mice after the first, second, and third hours of the experiment.

With regard to the state of the mice, it is NOT correct to say that

(A) *After the first hour there are only 75 healthy mice*
(B) After the first hour there are 129 sick mice
(C) After two and a half hours there are more healthy mice than sick mice
(D) Between the second and third hour the number of sick mice increased by 6.25 percentage points

Figure C3. Math sample questions

Notes: Correct answers are in *italics*. The sample question from the pre-reform exam was obtained from a version of the ICFES testing agency's website that was archived in January 1997 (available at https://web.archive.org/web/19980418191357/http://acuario.icfes.gov.co/12/122/1222/12223/Tipos.html in October 2016). The sample question from the post-reform exam was obtained from a September 2008 ICFES report entitled "State Assessment Tests in Colombia" (*"Evaluación con Pruebas de Estado en Colombia"*) (available at http://www.ieia.com.mx/materialesreuniones/1aReunionInternacionaldeEvaluacion/PONENCIAS18Septiembre/ConferenciasMagnas/MargaritaPenaBorrero.pdf in October 2016).

Panel A. Pre-reform sample question

Assertion: The only factor that determined the abolition of slavery in Colombia in the mid-nineteenth century was the economy.

Reason: In the mid-nineteenth century the formation of regional markets and the development of agriculture in our country made it necessary to establish freedom of labor.

(A) If the assertion and reason are true and the reason is a correct explanation of the claim, fill in oval A
(B) If the assertion and reason are true, but the reason is not a correct explanation of the claim, fill in oval B
(C) If the assertion is true but the reason is a false proposition, fill in oval C
(D) *If the assertion is false but the reason is a true proposition, fill in oval D*
(E) If both assertion and reason are false propositions, fill in oval E

Panel B. Post-reform sample question

In South America, archaeological finds of pottery—used for food preparation and storage of grain—have been interpreted as evidence of the strengthening of agriculture between the Andean cultures before the Inca Empire. These findings are indicative of agricultural and sedentary cultures because

(A) They reflect the broad expanse of corn, cacao, and vegetables
(B) There are no findings of hunting weapons made of stone
(C) *Nomadic activities, in contrast, require little ceramic production*
(D) Large irrigation systems are part of the same findings

Figure C4. Social sciences sample questions

Notes: Correct answers are in *italics*. The sample question from the pre-reform exam was obtained from a version of the ICFES testing agency's website that was archived in January 1997 (available at https://web.archive.org/web/19980418191357/http://acuario.icfes.gov.co/12/122/1222/12223/Tipos.html in October 2016). The sample question from the post-reform exam was obtained from a September 2008 ICFES report entitled "State Assessment Tests in Colombia" (*"Evaluación con Pruebas de Estado en Colombia"*) (available at http://www.ieia.com.mx/materialesreuniones/1aReunionInternacionaldeEvaluacion/PONENCIAS18Septiembre/ConferenciasMagnas/MargaritaPenaBorrero.pdf in October 2016).