Essays in the Economics of Labor and Higher Education

Evan Riehl

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Essays in the Economics of Labor and Higher Education

Evan Riehl

This dissertation examines the role of information in influencing both individuals' college outcomes and the productivity of a higher education system. It focuses in particular on large-scale educational reforms that raise different mechanisms than those in the existing literature on the returns to college attendance and college quality.

Recent work has shown that the choices of *whether* to attend college and *which* college to attend can both affect individuals' future earnings. These papers typically focus on a narrow subset of students or schools to credibly identify the effects of college choice. This dissertation instead uses data on the near universe of college students in an entire country to explore informational mechanisms that are difficult to isolate in existing work. To do this, I exploit reforms to the higher education system in Colombia that affect the information on individual ability that is transmitted to colleges, to employers, or to students themselves. This allows me to adapt traditional labor economic topics like employer learning (Jovanovic, 1979) and assortative matching (Becker, 1973) to the context of higher education. In addition, the large-scale nature of these reforms raises general equilibrium issues that may not arise from marginal changes in college admissions (e.g., Heckman, Lochner and Taber, 1998).

In Chapter 1, "Assortative Matching and Complementarity in College Markets," I examine one type of assortative matching in college markets: students with high socioeconomic status (SES) are more likely to attend high quality colleges. Assortativity matters if SES and college quality are complementary educational inputs. I develop an econometric framework that provides tests for the existence and sign of this complementarity. I implement these tests by exploiting a 2000 reform of the national college admission exam in Colombia, which caused a market-wide reduction in assortative matching in some regions of the country. I find that the reform lowered average graduation rates and post-college earnings in affected regions, consistent with a positive complementarity between SES and college quality. I also

find evidence of mismatch: part of these negative effects came from the low SES students who were shifted into higher quality colleges. However, both the market-wide and mismatch effects die out several cohorts after the exam reform, which suggests that complementarity may evolve with large-scale changes in assortativity.

In Chapter 2, "The Big Sort: College Reputation and Labor Market Outcomes," W. Bentley MacLeod, Juan E. Saavedra, Miguel Urquiola, and I ask how college reputation affects the process by which students choose colleges and find their first jobs. We incorporate a simple definition of college reputation—graduates' mean admission scores—into a competitive labor market model. This generates a clear prediction: if employers use reputation to set wages, then the introduction of a new measure of individual skill will decrease the return to reputation. We confirm this prediction by exploiting a natural experiment from the introduction of a college *exit* exam in the country of Colombia. Finally, we show that college reputation is positively correlated with graduates' earnings growth, suggesting that reputation matters beyond signaling individual skill.

Finally, in Chapter 3, "Time Gaps in Academic Careers," I ask if interruptions in students' academic careers can lower their overall schooling attainment. I study an academic calendar shift in Colombia that created a one semester time gap between high school and potential college entry. This brief gap reduced college enrollment rates relative to unaffected regions. Low SES students were more likely to forgo college, and individuals who did enroll after the gap chose higher paying majors. Thus academic time gaps can affect both the mean and the distribution of schooling attainment, with implications for the design of education systems and for wage inequality.

Contents

ACKNOWLEDGEMENTS

I am greatly indebted to W. Bentley MacLeod, Miikka Rokkanen, Juan E. Saavedra, and Miguel Urquiola for guidance and support throughout my graduate studies. For invaluable comments on my dissertation I thank Joseph Altonji, Peter Arcidiacono, Peter Bergman, Nicolás De Roux, Christopher Hansman, Adam Kapor, Nandita Krishnaswamy, Costas Meghir, Michael Mueller-Smith, Phil Oreopoulos, Kiki Pop-Eleches, Jonah Rockoff, Judith Scott-Clayton, Eric Verhoogen, Russell Weinstein, and participants in the Columbia University student colloquia. For help with the data I am grateful to Julian Mariño and Adriana Molina at the Colombian Institute for Educational Evaluation (ICFES), Luz Emilse Rincón at the Ministry of Social Protection, and Luis Omar Herrera at the Ministry of Education. All errors are my own.

For my parents, Ed and Sarah, and my sister, Emily, who fostered my love of education.

Chapter 1. Assortative Matching and Complementarity in College Markets

1. Introduction

College admission exam scores are often strongly correlated with socioeconomic status (SES). This contributes to a form of *assortative matching* in college markets—high SES students are more likely to attend high quality schools. Testing agencies in several countries have rewritten exams to decrease the correlation of scores with SES in an attempt to reduce assortativity in college admissions. For instance, this was a primary motivation for the 2016 overhaul of the SAT in the U.S. How do reductions in assortative matching from such reforms affect students' aggregate outcomes?

This paper shows that assortativity matters for aggregate outcomes if there is a *complementarity* between SES and college quality. A complementarity exists if, for example, top colleges have a comparative advantage in educating high SES students. Current research on college quality does not cleanly identify this complementarity, either because it cannot isolate mechanisms or it makes strong assumptions on unobservables. I develop an econometric framework to derive tests for the existence and sign of this complementarity in the presence of an exogenous shift in assortativity. I implement these tests by exploiting an admission exam reform in Colombia that reduced assortativity in some regions of the country. I find that the reform lowered graduation rates and post-college earnings—both region wide and for low SES students in particular—consistent with a positive complementarity between SES and college quality. However, these negative effects die out several cohorts after the reform, suggesting that complementarity may evolve with large-scale changes in assortativity.

My framework begins by defining two elements of the match between students and colleges: assortativity and complementarity. Assortativity is a feature of the distribution of matches. I define assortativity as the correlation coefficient between measures of student SES and college quality. A complementarity exists if student and college traits interact in the production of graduation or earnings outcomes. Formally, I say that the potential outcome function for students' human capital accumulation exhibits a complementarity if it is *not* additively separable in individual and college characteristics.

The purpose of my framework is to show that an admission exam reform that alters assortativity is informative about complementarity in college markets. I consider a reform that reduces the correlation of exam scores with SES. If this exam plays an important role in college admissions, then it should also reduce assortativity; low SES students gain access to higher quality colleges, and high SES students are displaced to lower-ranked schools.

I derive two propositions on the nature of complementarity between students and colleges. First, I show that the reduction in assortativity provides a clean test for the existence of a complementarity. The shift in assortativity allows me to estimate the aggregate effect of a market-wide reallocation of college quality, or what Graham (2011) calls an average reallocation effect. My first proposition states the average reallocation effect is zero unless there is a complementarity between students and colleges. Intuitively, there may be "winners" and "losers" from the reform, but on average these effects net out if there is no complementarity.

The test in Proposition 1 has advantages in identifying complementarity relative to work on the impact of admission to selective colleges (e.g., Dale and Krueger, 2002*a*; Hoekstra, 2009*a*). Selectivity papers typically cannot isolate the mechanisms through which colleges affect students' outcomes because they focus on a narrow set of individuals or schools to achieve credible identification. My framework shows that selectivity estimates are consistent with a wide range of models of the college production function, including those with no complementarity (e.g., linear-in-mean peer effects or student-invariant value added). Even when researchers examine heterogeneity by SES, variation in selectivity estimates may reflect differences in the colleges chosen by non-admitted students rather than a true complementarity. By contrast, the exam reform allows me to measure college quality effects across an entire market, which can separate models with and without a complementarity.

My second proposition shows how a change in assortativity can be used to test a benchmark assumption in research on college mismatch (Arcidiacono and Lovenheim, 2016). The mismatch hypothesis states that students who are academically unprepared for top colleges are better off attending lower-ranked schools. The best evidence of mismatch relies on the strong assumption that student/college matches are random conditional on observable traits

(Arcidiacono, Aucejo and Hotz, 2016). This allows researchers to uncover the causal returns to college quality from observed matches and then use these returns to explore counterfactual match outcomes. I can test this assumption in my context by comparing effects predicted by the mismatch methodology to those actually realized after the exam reform.

Proposition 2 states that under this assumption, the exam reform that I consider should reveal a positive market-wide complementarity, but it should not lead to mismatch. I show that pre-reform returns to college quality in my context are larger for high SES students than for low SES individuals, which suggests a positive complementarity. But these returns are positive even for low SES students, suggesting no mismatch. Thus if pre-reform data reveal the true returns to college quality, a decline in assortativity should 1) reduce market-wide graduation and earnings outcomes, but 2) increase these outcomes for low SES students.

The framework concludes by describing how these effects may evolve over time if students or colleges respond to the exam reform. An assumption underlying both propositions is that the distributions of student and college traits are stable across cohorts. This may not hold if, for example, colleges adapt their curricula to their new student bodies. Such responses are potentially important given that I consider larger changes in colleges' class composition than in the work noted above. I show that these responses may cause the predicted effects in Propositions 1 and 2 to differ between the first few cohorts exposed to the redesigned exam and students in later cohorts.

The empirical part of the paper uses administrative data from Colombia and a reform of its national admission exam to test the implications of Propositions 1 and 2. In 2000, this exam underwent a major redesign similar in spirit to the 2016 SAT overhaul. The new exam aimed to reduce socioeconomic bias by testing "competencies" rather than "content." I explore the effects of this reform using system-wide data that link admission scores, college records, and earnings for individuals in the two cohorts before and after the overhaul (1998–2001).

I begin by showing that the reform led to a dramatic reduction in the correlation of exam scores with SES. I define an index of socioeconomic traits that includes mother's education, high school quality, and gender, and I show that the correlation of this index with exam scores decreased sharply in the post-reform cohorts. Low SES students scored substantially higher on the new exam, and high SES students scored correspondingly lower. In several subjects, the decrease in the SES/exam score correlation was as large as 50 percent.

Next, I exploit features of the Colombian college system to identify effects of the exam reform on assortativity. Each administrative region has a large flagship public university that is often its most selective college. Flagship admissions affect the composition of students at nearby schools because college attendance is highly localized. The majority of flagship universities admit students solely based on scores from the national test, but some flagships administer their own admission exams. The reform was *ex ante* more likely to affect college admissions in regions where the flagship university used only national exam scores. The reform was likely to have fewer effects in regions where flagships used their own exams.

My "first stage" result confirms that the reform reduced student/college assortativity in regions where the national exam was more important for admissions. Specifically, the reform decreased the correlation of college quality—defined as school mean pre-reform admission scores—with the student socioeconomic index in affected regions relative to unaffected regions. Low SES students became more likely to attend high quality colleges relative to their control region peers, and high SES students were shifted into lower quality colleges.

I then present my main result: the reduction in assortativity led to a market-wide decrease in graduation rates and post-college earnings, suggesting a positive complementarity between SES and college quality. Overall graduation rates in affected regions fell by 1.4 percentage points relative to unaffected regions. Region-wide formal sector earnings, measured ten years after the admission exam, also fell by about one percent. These results hold in differences in differences regressions and in event-study graphs, which show sharp decreases in outcomes in the first cohort after the reform. This rejects models that do not exhibit a student/college complementarity, as stated in Proposition 1. Further, the decrease in market-wide outcomes is consistent with a positive complementarity, as predicted by Proposition 2.

However, I also find evidence of mismatch, which is inconsistent with the "conditional on observables" assumption in Proposition 2. I find that the market-wide reduction in graduation rates and earnings is partially driven by the lowest SES students, whose college quality *increased* with the reform. The low SES result is evidence of mismatch, but it is not what one would predict given the positive pre-reform return to college quality for these students. Together with the market-wide effects, the Colombian context thus provides mixed support for the benchmark assumption in mismatch research.

In the final section of the paper, I show that the negative graduation and earnings effects die out several cohorts after the exam reform. The above results are from the first two post-reform cohorts (2000–2001), which provide the cleanest identification and data coverage. But the question of whether the initial effects persist into later cohorts has important policy implications. I first show that the reform's effects on admission scores and on assortativity are stable over the next five cohorts (2002–2006); low SES students continue to score higher on the new exam, and in affected regions they continue to enroll in higher quality colleges. However, the negative graduation and earnings effects disappear by the sixth cohort after the reform (2005).[1] In other words, after several cohorts there are no longer differences in market-wide or low SES outcomes between affected and unaffected regions.

In sum, the Colombian reform provides evidence of a positive complementarity between SES and college quality, and it shows that some students may have better outcomes at lower-ranked schools. However, the results from more recent cohorts suggest that mismatch and complementarity are not fixed, inevitable concepts. There may be adjustment costs to admission reform, but the nature of college quality—and thus the productivity of a college system—may adapt to large-scale changes in the assignment of students to schools.

My paper adds to the mismatch literature by showing evidence of mismatch under weaker assumptions. Previous research has characterized the necessary conditions for mismatch (Arcidiacono et al., 2011) and the large gaps in academic preparation within colleges (Hoxby and Avery, 2013*b*; Dillon and Smith, 2016), but the few papers that document mismatch require strong assumptions on the distributions of unobservable traits (Arcidiacono et al.,

[1] These graduation and earnings effects are measured at younger ages due to data constraints.

2014; Arcidiacono, Aucejo and Hotz, 2016).[2] Further, I show that mismatch can affect labor market earnings in addition to the graduation outcomes examined in these papers. However, the short-term nature of my findings suggests that mismatch may not be solely due to class rank (Murphy and Weinhardt, 2014). Instead, mismatch may arise because colleges specialize in educating certain types of student, and these specializations may evolve in response to changes in a school's student body (Duflo, Dupas and Kremer, 2011).

My paper contributes to the selectivity literature (Dale and Krueger, 2002*a*, 2014*a*; Hoekstra, 2009*a*; Saavedra, 2009; Ketel et al., 2012; Hastings, Neilson and Zimmerman, 2013*b*; Goodman, Hurwitz and Smith, 2014; Canaan and Mouganie, 2015; Kirkebøen, Leuven and Mogstad, 2016) by showing that some mechanisms consistent with their findings—including linear peer effects or a fixed value added—cannot fully explain how colleges affect student outcomes. In contrast to standard results from this literature, I find that low SES students experience a negative return to college quality in the initial post-reform cohorts, and a zero return in later cohorts. My results suggest the conclusions from selectivity research may not apply to policies that cause large changes in the distribution of student/college matches.

My paper also relates to work the econometrics of assignment problems (Bhattacharya, 2009; Graham, 2011; Graham, Imbens and Ridder, 2014, 2016).[3] These models explicitly rely on the existence of a complementarity to generate non-zero reallocation effects under counterfactual match distributions. My framework differs in that it develops a reduced-form test for the existence of a complementarity. I also explore validity of the benchmark assumption that allows for counterfactual estimation in this literature. I find that pre-treatment data partially but incompletely explain the post-treatment effects, which highlights the difficulty of predicting counterfactual outcomes in contexts with social spillovers (Graham, Imbens and Ridder, 2010; Carrell, Sacerdote and West, 2013; Manski, 2013).[4]

[2] Several studies take reduced form approaches by exploiting affirmative actions bans (Cortes, 2010; Backes, 2012; Hinrichs, 2012, 2014), but they find no direct evidence of mismatch.

[3] These papers focus on contexts where the output of a match is observed by the econometrician. Other work explores contexts in which one observes the match distribution but not match output (Choo and Siow, 2006*a*,*b*). See Chiappori and Salanie (2016) for a review of this literature.

[4] However, I present evidence that non-linear peer effects (Tincani, 2014; Booij, Leuven and Oosterbeek, 2015) cannot fully explain the difference between the predicted and realized effects in my context.

More broadly, my paper contributes to a large theoretical and empirical literature on matching models. Since Becker (1973)'s canonical study of assortative mating, many papers have asked whether the match function in different markets is supermodular, which implies a positive complementarity between two populations (Topkis, 1998; Siow, 2015). My paper provides evidence of a supermodular but malleable match function in the context of college markets. This relates to work on numerous educational matching mechanisms, including centralized assignment (Gale and Shapley, 1962; Abdulkadiroğlu, Pathak and Roth, 2005), affirmative action (Bertrand, Hanna and Mullainathan, 2010; Antonovics and Backes, 2014; Bagde, Epple and Taylor, 2016), "percent plans" like the Texas Top Ten Percent rule (Cullen, Long and Reback, 2013; Black, Cortes and Lincove, 2015; Kapor, 2015), and admission tests like the SAT (Rothstein, 2004).

The paper proceeds as follows. Section 2 develops the framework, and Section 3 describes the Colombian exam reform. Section 4 shows that the reform reduced assortative matching, and Section 5 presents its effects on human capital accumulation. Section 6 examines the evolution of these effects beyond the initial post-reform cohorts. Section 7 concludes.

2. Econometric framework

This section presents a framework that uses an exogenous change in assortative matching in a college market to test for a complementarity between student socioeconomic status (SES) and school quality. I begin by defining student/college characteristics and the concepts of assortativity and complementarity. I then consider the effects of an admission exam reform that reduces student/college assortativity. I show how this reform can be used to test for the existence of a complementarity, and I discuss how this analysis contributes to existing work on college selectivity. Next, I show how the sign of this complementarity that is revealed by the exam reform can be used to test the key "conditional on observables" assumption in the literature on college mismatch. Finally, I discuss potential student and college responses to the reform that could cause this complementarity to evolve over time.

The setup of this framework is a simplified version of that in Graham (2011), which is motivated by the study of *reallocation effects*—changes in an outcome that arise from alternate matchings of two populations. This typically requires strong assumptions on the distributions of unobservable characteristics or on the form of the matching function. My framework differs in that it exploits an admission exam reform for a reduced form exploration of student/college reallocation effects.

2.1. **Students, colleges, and human capital.** This framework considers the matching of two populations in a market—students and colleges. It begins by defining the characteristics of both populations that affect students' human capital accumulation.

The first population consists of students, who have academic ability that can be decomposed into observable and unobservable components. Students are indexed by i and have ability A_i, which is given by:

$$A_i = X_i + \epsilon_i. \tag{1}$$

X_i is an index of an individual's observable socioeconomic characteristics. In the empirical analysis below I define X_i to be a socioeconomic index based on mother's education, high school quality, and gender; specifically, I use the mean admission exam score within cells defined by these characteristics. Thus X_i gives a finely-grained measure of an individual's expected performance on the admission exam given her socioeconomic characteristics. I normalize X_i to have mean zero and standard deviation one. The residual term ϵ_i reflects unobservable components of an individual's academic ability that are independent of X_i. Thus ϵ_i captures those aspects of individuals' abilities that are uncorrelated with their observable socioeconomic backgrounds.

The second population consists of colleges, which also have observable and unobservable characteristics. Colleges are indexed by c, and each college has quality Q_c that can be decomposed in an analogous manner:

$$Q_c = W_c + \nu_c. \tag{2}$$

W_c is an observable measure of college quality, which is normalized to mean zero and standard deviation one. Below I define W_c to be the average admission exam score of a college's students. Thus W_c gives an individual's expected performance on the admission exam given her college. I denote the unobservable component of college quality by ν_c, which I define to be independent of W_c. This means that W_c will reflect other aspects of a college's quality that are correlated with the average exam score of its students, including its financial resources and faculty expertise. W_c may also be correlated with other unobservable characteristics of a college's students, such as their family resources. ν_c thus reflects all other aspects of a college's financial, faculty, or peer quality that are uncorrelated with mean exam scores.

Both individual and college characteristics can affect a student's human capital accumulation.[5] This notion is captured by the potential match output function:

$$Y_{(i,c)} = Y_{(i,c)}(A_i, Q_c). \tag{3}$$

$Y_{(i,c)}$ is the human capital that would result from the potential assignment of student i to college c, as measured by graduation outcomes or subsequent labor market earnings. This output depends on a student's academic ability, A_i, and the quality of her college, Q_c. Potential output is defined for all combinations of these characteristics, but the data contain only one realization of output, Y_{ic}, based on the college a student actually attended.

2.2. **Assortativity and complementarity.** This paper highlights two features of the match between students and colleges: assortativity and complementarity.

Assortativity is a feature of the distribution of student/college matches. In particular, positive assortative matching between students and colleges occurs when high socioeconomic students are more likely to attend high quality colleges. In matching research, assortativity is frequently measured using the correlation coefficient between the characteristics of two populations. I follow this convention in defining assortativity in college markets:

[5] Throughout this paper I use the phrase "human capital" to refer to graduation and post-college earnings outcomes. Colleges may affect these outcomes through channels other than pure human capital accumulation. For example, the college a student attends may affect her earnings through a signaling mechanism (MacLeod et al., 2015). My use of the term "human capital" is not meant to rule out these other mechanisms.

Definition. The *assortativity* of student/college matches in a market is the correlation coefficient, ρ, between the socioeconomic index, X_i, and observable college quality, W_c.

Since X_i and W_c are both defined to have standard deviation one, the correlation coefficient between these two characteristics is equivalent to the ρ coefficient in the linear regression:[6]

$$W_c = \rho X_i + u_{ic}, \tag{4}$$

where u_{ic} is the residual. Equation (4) yields a convenient representation of ρ for the empirical analysis below.

While assortativity is a feature of the match distribution, complementarity is a characteristic of the potential match output function, $Y_{(i,c)}$. A complementarity exists if there is an interactive effect of individual ability and college quality in the production of human capital. Formally, I define complementarity as follows:[7]

Definition. The potential match output function, $Y_{(i,c)}$, exhibits a *complementarity* if is not additively separable in A_i and Q_c.

For example, a simple potential output function with no complementarity is $Y_{(i,c)} = \theta^a A_i + \theta^q Q_c$. By contrast, the potential output function $Y_{(i,c)} = \theta A_i Q_c$ exhibits a complementarity; the return to college quality, Q_c, is higher for students with higher academic ability, A_i.

Note that I define assortativity in terms of the observable indices X_i and W_c, which allows me to compute a value of ρ from the data. By contrast, I define complementarity using A_i and Q_c, which include both observable and unobservable components. Below I show that a change in assortativity provides a clean test of complementarity defined in terms of A_i and Q_c. I also show that stronger assumptions are necessary to interpret this complementarity as one between the observable characteristics X_i and W_c.

---

[6] Specifically, $\rho = cov(X_i, W_c)/\sigma_X \sigma_W = cov(X_i, W_c)/var(X_i)$ since $\sigma_X = \sigma_W = var(X_i) = 1$. I also omit a constant term from equation (4) because both X_i and W_c are mean zero.

[7] Following Graham (2011), I use the word "complementarity" to denote either a positive or a negative interactive effect. Some researchers use "complementarity" for positive interactions and "substitutability" for negative interactions.

2.3. **Admission exam reform and a reduction in assortativity.** This framework considers a college admission exam reform that reduces the correlation of exam scores with student SES, and thus reduces the assortativity of student/college matches.

Let τ_{it} be the score of individual i on a college admission exam in period t. In this framework I consider two periods defined by individuals' exam cohorts, $t \in \{0, 1\}$, where 0 represents the cohort before the admission exam reform, and 1 is the post-reform exam cohort. I use a linear specification to define the relationship between test scores and observable socioeconomic characteristics in each cohort:

$$\tau_{it} = \rho_t^{exam} X_i + u_{it}. \tag{5}$$

As above, I normalize exam scores, τ_{it}, to be mean zero and standard deviation one; thus ρ_t^{exam} represents both the linear regression coefficient and the correlation coefficient between exam scores and the socioeconomic index in cohort t.

An admission exam reform that reduces the correlation of test scores with students' socioeconomic backgrounds implies that:

$$\rho_1^{exam} < \rho_0^{exam}. \tag{6}$$

In other words, the correlation between test scores, τ_{it}, and the socioeconomic index, X_i, is lower in the post-reform cohort. Equation (6) gives the key characteristic of the admission reform instrument that I document empirically in Section 3.

The reduction in the correlation of X_i with exam scores in equation (6) does not perfectly translate into an effect on college admissions because tuition costs and other preferences also influence students' college choices. Nonetheless, if exam scores play an important role in admissions, the reform should also reduce the assortativity of student/college matches. Allowing the definition of assortativity in equation (4) to differ between the two exam cohorts, this implies:

$$\rho_1 < \rho_0. \tag{7}$$

Equation (7) is the "first-stage" prediction for my empirical analysis, which I test in Section 4.

2.4. **Existence of complementarity and the college selectivity literature.** This section shows that a change in assortativity from admission exam reform provides evidence on the existence of a complementarity between student and college characteristics. I then show that the estimand from such a reform has advantages in identifying a complementarity relative to empirical designs that are common in the literature on college selectivity.

The decline in assortativity represented by equation (7) implies a reallocation of student/college matches across an entire market. This allows me to estimate the market-wide effect of this new match distribution on human capital accumulation. This estimand is what Graham (2011) calls the *average reallocation effect*, and it is given by:

$$\beta = E_1[Y_{ic}] - E_0[Y_{ic}], \tag{8}$$

where $E_0[Y_{ic}]$ and $E_1[Y_{ic}]$ denote the average values of an observable measure of human capital, Y_{ic}, in the cohorts before and after the exam reform. The average reallocation effect, β, is thus the change in market-level match output from the exam overhaul. It is the primary estimand of this paper, although in the empirical analysis below I estimate a differences in differences version of β.[8] In this framework I use the single difference estimand (8) for simplicity of exposition.

The key assumption necessary to identify a complementarity is that the distributions of academic ability and college quality are stable over time.[9] In the case of academic ability, this assumption means that the distributions of both the observable, X_i, and unobservable, ϵ_i, components of A_i are the same in the pre- and post-reform cohorts. In my analysis,

[8] The differences in differences specification uses regional variation in the effects of an exam reform to net out concurrent common shocks to human capital accumulation. Specifically, I define treated (T) and control (C) regions of Colombia based on the potential effects of the exam reform, and estimate the average reallocation effect as $\beta = (E_1^T[Y_{ic}] - E_0^T[Y_{ic}]) - (E_1^C[Y_{ic}] - E_0^C[Y_{ic}])$. See Section 4 for details on the differences in differences analysis.

[9] In the differences in differences specification below, the assumption is that there are parallel trends in these distributions across regions.

X_i includes only pre-determined characteristics that cannot be affected by the exam reform. The assumption for ϵ_i is stronger because this term includes aspects of individual ability that are potentially influenced by the reform. For example, the overhaul could induce students to study harder for the new exam, raising their human capital.

The assumption of stability in college quality, Q_c, means that the distributions of both the observable, W_c, and unobservable, ν_c, components are the same in the pre- and post-reform cohorts. Below I define W_c using only pre-reform characteristics, and any changes in its distribution are testable. The stability of unobservable college quality, ν_c, is the strongest of these assumptions. For example, colleges may change curricula or support services in response to the reform. More importantly, ν_c includes aspects of a student's college peers, which necessarily change as a result of the decline in assortativity from the exam reform. Thus if peer effects are an important component of college quality, the distribution of ν_c may not be stable across exam cohorts.

I examine the empirical evidence related to these competing explanations in Section 4.3. I argue that the pattern of results in the initial cohorts after the reform is difficult to explain through changes in the distributions of unobservable ability, college quality, or peer effects. However, in subsequent cohorts I present evidence of student/college responses that alter the reform's initial human capital effects.

If the distributions of academic ability, A_i, and college quality, Q_c, are stable across exam cohorts, it immediately follows that the average reallocation effect (8) provides a clean test of the existence of a complementarity. Any potential output function that does not exhibit a complementarity is additively separable in A_i and Q_c, which means it can be written in the form $Y_{(i,c)} = f(A_i) + g(Q_c)$ for some functions $f(\cdot)$ and $g(\cdot)$. The average reallocation effect under this potential output function is:

$$\begin{aligned}
\beta &= E_1[Y_{ic}] - E_0[Y_{ic}], \\
&= E_1[f(A_i) + g(Q_c)] - E_0[f(A_i) + g(Q_c)], \\
&= E_1[f(A_i)] - E_0[f(A_i)] + E_1[g(Q_c)] - E_0[g(Q_c)], \\
&= 0,
\end{aligned}$$

where the last line follows from the assumption that the academic ability and college quality distributions are stable. Thus the average reallocation effect of a decline in assortativity should be zero unless there is a complementarity between student and college characteristics. This is summarized in the following proposition:

Proposition 1. *If the distributions of student ability and college quality are stable over time, a decline in assortativity has no effect on market-level human capital accumulation unless student and college traits are complementary.*

The idea that a complementarity is necessary to generate a non-zero reallocation effect is not new. In particular, Graham (2011) notes that the study of assignment problems is only interesting in the presence of a complementarity. The contribution of Proposition 1 is to exploit an exogenous admission exam reform to provide a test for the existence of such a complementarity. By contrast, existing empirical work on reallocation effects typically measures complementarity using strong assumptions on the distributions of unobservable match characteristics (see, e.g., Bhattacharya, 2009; Graham, Imbens and Ridder, 2010), as I discuss in the next subsection.

The average reallocation effect in (8) has advantages in identifying complementarity relative to empirical designs that are common in another literature that takes a reduced form approach to measuring college quality effects—the "college selectivity" literature (e.g., Dale and Krueger, 2002*a*; Hoekstra, 2009*a*). The estimand of interest in selectivity studies takes the general form:

$$\beta^{CS} = E[Y_{ic}|A_i = a, Q_c = q_H] - E[Y_{ic}|A_i = a, Q_c = q_L]. \tag{9}$$

The college selectivity effect, β^{CS}, measures the human capital return to attending a high selectivity college with quality q_H rather than a less selective college with lower quality q_L. This return is estimated for students of a given academic ability, a, which typically corresponds to a marginally accepted student; for example, numerous studies use regression discontinuity designs with a given by the location of an admission threshold for the selective college.[10]

The first three rows of Table 1 show predictions for the college selectivity estimand in (9) and the average reallocation estimand in (8) under simple potential match output functions that do not exhibit a complementarity. The first row represents a model in which colleges contribute to individual outcomes through a value added component, θQ_c, that does not depend on any student characteristics. The second row defines a case in which an individual's human capital accumulation depends only on the mean peer ability at her college, $\theta \bar{A}_c$.[11] This linear-in-mean peer ability model is widely used in both theoretical (e.g., Epple and Romano, 1998*b*) and empirical (e.g., Manski, 1993) work on peer effects. The third row depicts a model in which a student's human capital accumulation depends only on her class rank. This is often a key part of the implicit or explicit model used in college mismatch papers (e.g., Arcidiacono and Lovenheim, 2016), which argue that students can become discouraged or receive poor grades if their academic ability is too far below that of their classmates.[12]

---

[10] Hoekstra (2009*a*), Saavedra (2009), Ketel et al. (2012), Hastings, Neilson and Zimmerman (2013*b*), Goodman, Hurwitz and Smith (2014), Canaan and Mouganie (2015), Kirkebøen, Leuven and Mogstad (2016) all use regression discontinuity thresholds for college admissions. Other papers use similar empirical designs in the context of high school admissions (e.g., Pop-Eleches and Urquiola, 2013; Abdulkadiroğlu, Angrist and Pathak, 2014).

[11] Thus in this model, college quality, Q_c, depends only on the mean ability at a college, $\bar{A}_c$.

[12] Several more complicated models of class rank, such as a student's percentile rank, yield similar predictions to this simple model.

Table 1. Predicted selectivity and reallocation effects from simple match functions

Description	Potential match output, $Y_{(i,c)}$	College selectivity effect, β^{CS}	Average reallocation effect, β
Common value added	θQ_c	$\theta(q_H - q_L)$	0
Linear-in-mean peer ability	$\theta \bar{A}_c$	$\theta(\bar{a}_H - \bar{a}_L)$	0
Class rank	$\theta(A_i - \bar{A}_c)$	$\theta(\bar{a}_L - \bar{a}_H)$	0
Complementarity	$\theta A_i Q_c$	$\theta a(q_H - q_L)$	$\theta(E_1[A_i Q_c] - E_0[A_i Q_c])$

Notes: The average reallocation effect, β, is defined in equation (8). The college selectivity effect, β^{CS}, is defined in equation (9).

For each of these three models, the college selectivity effect, β^{CS}, is potentially non-zero, and is given by the difference in quality of the high and lower selectivity colleges. For example, in the linear-in-mean peer ability model, the selectivity effect is $\theta(\bar{a}_H - \bar{a}_L)$, which reflects the difference in the mean student ability at the two colleges. By contrast, the average reallocation effect under all three models is zero. This arises because there are no interactive effects between student characteristics and college quality, and thus market-level output does not depend on which students are assigned to which colleges. In the peer ability model, for instance, students that are shifted into colleges with high mean ability after the reform may see gains in their outcomes, but these gains are offset by losses among students displaced to colleges with lower mean ability. In net, the reallocation of student/college matches has no effect on market-level output under such a model.

The last row of Table 1 depicts a potential output function with a simple complementarity between student and college characteristics, $Y_{(i,c)} = \theta A_i Q_c$. The college selectivity effect under this model is again non-zero, and it captures the quality difference between high and lower selectivity colleges for students with ability $A_i = a$. However, a decline in market-level assortative matching implies that the average value of $A_i Q_c$ differs before and after the exam reform. Thus, the average reallocation effect is also potentially non-zero.

Table 1 highlights the benefit of using the average reallocation estimand (8) in identifying complementarity. The average reallocation effect generates different predictions for models with and without a complementarity, while the college selectivity effect is potentially non-zero in all cases.

Furthermore, it is difficult to identify a complementarity even when researchers examine heterogeneity in college selectivity estimates across different types of students. Researchers frequently estimate the college selectivity effect, β^{CS}, for students who vary on some observable dimension, X_i, like race or family income; that is, they estimate $E[\beta^{CS}|X_i = x]$ for different values of x. For example, some papers find that college selectivity effects are larger for low-income students (Dale and Krueger, 2002*a*). This result might appear to suggest a negative complementarity between college quality and family income, but this need not be the case.

Suppose that, as in the first row of Table 1, potential match output is given by a common value added term, $Y_{(i,c)} = \theta Q_c$. Thus there is no complementarity between students and colleges. For all students, β^{CS} measures the effect of admission to the same selective college with quality q_H, but the counterfactual college that students attend if they are not admitted may vary across individuals. Thus the value added component for non-admitted students (q_L in Table 1) is actually an average of the value added terms for all colleges attended by these students. Heterogeneity in college selectivity effects is therefore given by:

$$E[\beta^{CS}|X_i = x] = \theta(q_H - \frac{1}{N_x} \sum_{i:X_i=x} q_{c_i}),$$

where q_{c_i} is quality of the college attended by student i, and N_x is the number of non-admitted students for whom $X_i = x$. Since $E[\beta^{CS}|X_i = x]$ depends on the counterfactual colleges chosen by non-admitted students, selectivity effects can vary across values of x even though there is no complementarity. For example, β^{CS} would be larger for low-income students if the schools they attend when they are not admitted to the selective college tend to have lower value added than those of high-income students.

If researchers can also control for the quality of the counterfactual college, q_L, then heterogeneity in college selectivity effects is more likely to reveal a true complementarity. With the notable exception of Kirkebøen, Leuven and Mogstad (2016), such empirical designs are rare in the selectivity literature. Yet even if counterfactual conditioning is possible, college selectivity estimates only reveal complementarity for marginally-admitted students and for the colleges included in the study. These effects may not reveal the average value of such a complementarity across an entire college market or in the presence of large-scale changes in the student/college match distribution, which is the focus of my analysis.

2.5. **Sign of complementarity and the college mismatch literature.** This section shows that a change in assortativity from admission exam reform can be used to test a key assumption in the literature on college mismatch: students and colleges are randomly matched conditional on observables. I further show that adopting this assumption yields the predictions that, in my context, a reduction in assortativity should decrease market-wide average graduation rates or earnings, but it should increase these outcomes for low SES students. In other words, under this assumption the data predict a positive complementarity between student SES and college quality but no mismatch for low SES individuals.

The mismatch hypothesis states that the return to college quality may be negative for students who are academically unprepared for top colleges (Arcidiacono and Lovenheim, 2016). The best evidence of mismatch comes from structural models that estimate the change in human capital accumulation that occurs in a counterfactual assignment of students to colleges. For example, Arcidiacono, Aucejo and Hotz (2016) show that overall science and technology graduation rates in the University of California system would have been higher had admissions been based more on measures of academic preparation—like GPA and SAT scores—and less on minority status. Related work also takes a structural approach to the counterfactual matching of two populations in other contexts, including roommate assignment (Bhattacharya, 2009) and segregation (Graham, Imbens and Ridder, 2010).

This work requires strong assumptions about the unobservable characteristics of the two populations under consideration. In this framework I use a stylized regression model to illustrate this assumption and highlight its key consequences for counterfactual outcomes. In the empirical analysis below I use non-parametric regressions that align more closely with the estimating equations in mismatch research.

To explore the effects of counterfactual student/college matches, researchers estimate an equation analogous to the following specification:

$$Y_{ic} = \alpha + \theta^x X_i + \theta^w W_c + \kappa X_i W_c + e_{ic}. \tag{10}$$

Equation (10) expresses human capital, Y_{ic}, as a parametric function of characteristics observable to the researcher: the student socioeconomic index, X_i, and college quality, W_c. This specification also includes a term, $X_i W_c$, that allows for a complementary effect of observable student and college traits. The parameter κ measures how the return to college quality, W_c, varies with an individual's socioeconomic background, X_i. The unobservable determinants of human capital, including individual ability, ϵ_i, and college quality, ν_c, are lumped into the residual, e_{ic}.

Equation (10) is a stylized representation of the observable components of potential output, but an advantage of this specification is that it yields a simple form for the average match output in a market:

$$\begin{aligned} E[Y_{ic}] &= \alpha + \theta^x E[X_i] + \theta^w E[W_c] + \kappa E[X_i W_c], \\ &= \alpha + \kappa\rho. \end{aligned} \tag{11}$$

The simplification in (11) comes from the fact that X_i and W_c are both mean zero and standard deviation one; thus, the $E[X_i]$ and $E[W_c]$ terms disappear, and $E[X_i W_c]$ is equal to the assortativity of student/college matches, ρ, as defined by equation (4).[13] Average output in a market, $E[Y_{ic}]$, is therefore given by the product of the complementarity, κ, and assortativity, ρ, of student and college characteristics (plus a constant term, α).

This specification also gives a convenient representation of the average reallocation effect—the change in average output from the exam reform given by equation (8). Using 0 and 1 subscripts to denote the values of each parameter in the cohorts before and after the exam

[13] In other words, $E[X_i] = E[W_c] = 0$, and $\rho = cov(X_i, W_c)/var(X_i) = E[X_i W_c] - E[X_i]E[W_c] = E[X_i W_c]$.

overhaul, the average reallocation effect can be written as:

$$\beta = E_1[Y_{ic}] - E_0[Y_{ic}]$$

$$= \kappa_1\rho_1 - \kappa_0\rho_0 \tag{12}$$

$$= \kappa_0(\rho_1 - \rho_0) + (\kappa_1 - \kappa_0)\rho_1. \tag{13}$$

Equation (12) shows that the average reallocation effect depends on the complementarity, κ, and assortativity, ρ, of student and college characteristics in the pre- and post-reform cohorts.[14] Equation (13) decomposes the average reallocation effect into two terms.[15] The first term, $\kappa_0(\rho_1 - \rho_0)$, represents the *distribution effect* of reallocation. This term measures the effect of a change in the assortativity of matches, ρ, given the initial value of student/college complementarity, κ_0. The distribution effect captures the essence of how researchers explore the effects of counterfactual matches. Researchers can estimate initial assortativity, ρ_0, and complementarity, κ_0, from the data, and the goal is to examine how outcomes changes under counterfactual distributions, ρ_1.

The second term, $(\kappa_1 - \kappa_0)\rho_1$, reflects a change in complementarity, κ, applied to the post-reform assortativity of matches, ρ_1. This term is a *selection bias* component because the complementarity that arises under the new allocation, κ_1, is unknown—it can depend on unobservable student/college match characteristics. To deal with this selection bias term, a benchmark assumption in structural work is that students and colleges are randomly matched conditional on their observable types, X_i and W_c. This assumption, which Graham (2008) calls *double randomization*, ensures that the joint distribution of unobservable student and college traits, ϵ_i and ν_c, is the same across all observed matches. For example, Arcidiacono, Aucejo and Hotz (2016) assume that student/college matches are random conditional on indicators for the sets of colleges to which students applied and were accepted—a strategy first proposed by Dale and Krueger (2002*a*). The assumption that matches are generated

[14] Note that the constant term, α, falls out because of the assumption of stability in the distributions of individual ability and college quality.

[15] Equation (13) is derived by adding and subtracting $\kappa_0\rho_1$ from equation (8).

randomly implies that κ_0 is the true, causal value of complementarity, and is therefore equal to κ_1. Thus the selection bias term $(\kappa_1 - \kappa_0)\rho_1$ disappears from (13), and researchers can estimate counterfactual reallocation effects by changing the match distribution ρ_1.

Adopting the double randomization assumption allows me to predict the consequences of an admission exam reform in my context. If there is no selection bias term, the sign of the average reallocation effect in (13) is equal to the sign of the distribution effect, $\kappa_0(\rho_1 - \rho_0)$. The exam reform reduces assortativity as given by assumption (7), so $(\rho_1 - \rho_0)$ is negative. It remains to estimate the sign of the pre-reform student/college complementarity, κ_0. One would observe $\kappa_0 > 0$ if high SES students benefit more from attending top colleges than students with lower values of X_i. Conversely, $\kappa_0 < 0$ could arise if low SES students benefit more from admission to top colleges.

Figure 1 shows evidence that $\kappa_0 > 0$ in the Colombian college market (see Sections 3 and 4 for more details on the context, data, and variable definitions). The horizontal axis depicts an index of students' socioeconomic status, X_i, and the vertical axis represents one measure of human capital accumulation, Y_{ic}: log earnings measured ten years after taking a college admission exam. The figure divides students into two groups based on the observed quality of their colleges, W_c. The solid black line contains students from above the median quality, while the light-red dashed line depicts students at colleges below the median of W_c. The fanning out of the two lines as X_i increases suggests that κ_0 is positive. Students of all socioeconomic backgrounds have higher earnings at top colleges than those at low quality colleges, but this gap is significantly larger for high SES students. Appendix A.1 shows that other measures of human capital accumulation, including drop-out and graduation rates, also exhibit a positive observed complementarity, although they are less pronounced.

If $\kappa_0 > 0$, then the distribution effect in equation (13), $\kappa_0(\rho_1 - \rho_0)$, is negative. Thus the overall effect of a reduction in assortativity is a decline in human capital accumulation unless the selection bias term is sufficiently positive to offset the distribution effect. The selection bias component can be non-zero if the change in student/college matches also affects sorting on the unobservable determinants of human capital, ϵ_i or ν_c. For example, hard-working

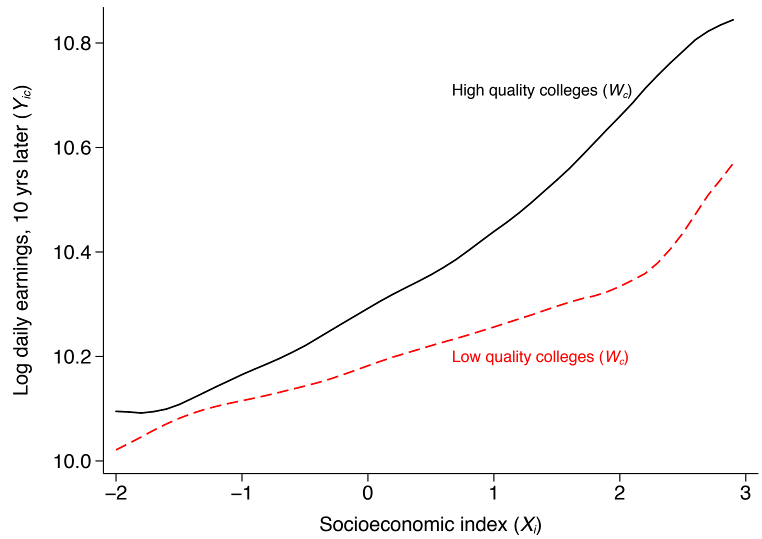


FIGURE 1. Observed complementarity in earnings

Notes: This figure uses data on college enrollees who took the Colombian national admission exam in 1998–1999. See Sections 3 and 4 for details on the data and variable definitions. The solid black line depicts the non-parametric relationship between a socioeconomic index, X_i, and log earnings measured 10 years after taking the admission exam, Y_{ic}, for students who attended colleges with quality, W_c, above the median. The light-red dashed line depicts the analogous relation for students at colleges below the median of W_c.

students may learn more at top colleges, and the exam reform may change the relationship between student work ethic and college quality. If work ethic is unobservable but correlated with X_i, this could affect estimates of κ. A reduction in assortative matching may also affect κ through changes in the peer characteristics at a student's college, which are not captured in the simple interaction term X_iW_c.

A similar derivation can be used to predict the reform's effects for students with different levels of SES. Specifically, define the average reallocation effect for individuals with a value of the socioeconomic index $X_i = x$ as:

$$\beta^x = E_1[Y_{ic}|X_i = x] - E_0[Y_{ic}|X_i = x]. \tag{14}$$

Since X_i is a pre-determined characteristic, the only assumption necessary to interpret this estimand as the causal effect of the exam reform is stability in the distribution of X_i over time. β^x thus gives the reduced-form effect of the decline in assortative matching for different values of X_i.

However, one can use the regression model (10) and a "conditional on observables" assumption analogous to double randomization to derive a prediction for the value of β^x:

$$\begin{aligned}
\beta^x =& E_1[\alpha + \theta^x X_i + \theta^w W_c + \kappa X_i W_c + e_{ic} | X_i = x] - \\
& E_0[\alpha + \theta^x X_i + \theta^w W_c + \kappa X_i W_c + e_{ic} | X_i = x] \\
=& (\theta^w + \kappa x)\Big(E_1[W_c | X_i = x] - E_0[W_c | X_i = x]\Big) \qquad (15)
\end{aligned}$$

The derivation of equation (15) uses two related assumptions. The first assumption is $E_1[e_{ic}|X_i = x] - E_0[e_{ic}|X_i = x] = 0$, which implies that the determinants of human capital accumulation other than the observable characteristics X_i and W_c have the same mean in the pre- and post-reform cohorts. The second assumption is that the values of the parameters θ^w and κ are the same in the two exam cohorts. This is analogous to double randomization in that it assumes that conditioning on X_i and W_c allows the researcher to uncover the true value of these parameters.[16]

Given these assumptions, equation (15) defines β^x as the coefficient $(\theta^w + \kappa x)$ times the difference in observable college quality, W_c, for individuals with $X_i = x$. The decline in assortativity given by equation (7) implies that $E_1[W_c|X_i = x] - E_0[W_c|X_i = x]$ is positive for students with low values of x, and it is negative for high values of x. In other words, low SES students experience increases in college quality with the exam reform, while high SES students are shifted into lower quality colleges.

The sign of β^x thus depends on the sign of the coefficient $(\theta^w + \kappa x)$, which includes the college quality effect, θ^w, the complementarity coefficient, κ, and the value of the socioeconomic index, x. These terms come from the stylized regression model (10), but they capture the basic intuition in the mismatch literature of a tradeoff between between college quality, θ^w, and academic fit, κx (Arcidiacono and Lovenheim, 2016). The college quality component is typically hypothesized to be positive, and thus if θ^w is large enough then there is no mismatch—all students, regardless of their socioeconomic backgrounds—benefit from

[16] Equation (15) also uses the stability assumption $E_1[\alpha + \theta^x X_i | X_i = x] - E_0[\alpha + \theta^x X_i | X_i = x] = 0$.

attending higher quality colleges. Conversely, mismatch arises for low SES students if the fit component, κx, is larger in magnitude than the college quality component, θ^w.

Figure 1 suggests that in the Colombian context, the coefficient $(\theta^w + \kappa x)$ is positive for all values of the socioeconomic index, X_i. The solid black line, which represents average earnings at high quality colleges, is always above the red dashed line, which gives average earnings at low quality colleges. Appendix A.1 shows that this pattern also holds for drop-out and graduation outcomes. Thus, if the double randomization assumption holds in Colombia, the sign of the reallocation effect for individuals with $X_i = x$ is given by the sign of the term $E_1[W_c|X_i = x] - E_0[W_c|X_i = x]$. This suggests that the admission exam reform should increase human capital accumulation for low SES students (i.e., $\beta^{x_L} > 0$), who experience increases in the college quality on average. Conversely, the reform should reduce human capital accumulation for high SES students (i.e., $\beta^{x_H} < 0$), whose average college quality decreases in the post-reform cohort.

Proposition 2 summarizes the market-wide and individual-level predictions on the sign of complementarity given the double randomization assumption.

Proposition 2. *If the double randomization assumption holds in the Colombian context, then a decline in the assortativity of student/college matches ($\rho_1 < \rho_0$) should:*

- *Decrease market-wide human capital accumulation ($\beta < 0$),*
- *Increase human capital outcomes for low SES students ($\beta^{x_L} > 0$), and*
- *Decrease human capital outcomes for high SES students ($\beta^{x_H} < 0$).*

In sum, an admission exam reform provides an opportunity to test the double randomization in the Colombian context. Estimates of the average reallocation effect, which include both the distribution and selection bias effects from equation (13), can be compared to the predicted signs from the distribution effects alone. The test is particularly compelling if there is exogenous variation in the effects of the admission reform, which I document in my setting in Sections 3 and 4. By exploiting the variation in student/college matches from an exam overhaul reform, the principal assumptions for this paper are weaker than those common in structural work on college mismatch.

The above discussion used a parametric model to characterize reallocation effects. In the empirical section below, I use non-parametric specifications to test Proposition 2. Instead of the assortativity (ρ) and complementarity (κ) parameters, I calculate non-parametric versions of the distribution of student/college matches ($Pr[X_i = x, W_c = w]$) and match function ($E[Y_{ic}|X_i = x, W_c = w]$). This specification is closer to those used in the mismatch literature.

2.6. **Longer-term responses to the reform.** This section discusses three potential ways in which students or colleges could respond to the exam reform. Each possibility would cause outcomes observed in the initial cohorts exposed to the new exam to differ from those in later cohorts.

First, over time students may respond to the reform in a way that undoes some of its effects on the distribution of scores. For example, high SES students may engage in tutoring that improves their performance on the new exam. Students may also begin to select high schools with better performance on the new exam. Such affects could alter the correlation between the socioeconomic index, X_i, and students' exam scores, τ_{it}, as defined by the parameter ρ^{exam}. In other words, the strength of the exam reform instrument—captured by equation (6)—may evolve across post-reform cohorts.

Second, students' college choices may change over time even conditional on exam scores. Students with the same scores as their peers in previous cohorts may end up in different colleges if schools adjust admission standards in response to their changing student bodies, or if students' preferences change in the response to the exam overhaul. This would lead to a change in the assortativity parameter, ρ, and it would cause the "first stage" prediction (7) to differ between initial and subsequent cohorts.

Third, even if the reform had a lasting effect on students' exam scores and college choices, the human capital effects of the reform can be temporary if the distributions of student ability or college quality respond to the reform. Both Propositions 1 and 2 rely on the assumption that the distributions of student ability and college quality are stable over time. This may not hold if, for example, colleges respond to their new student bodies by increasing support services for low SES students, or by adjusting the content or target levels of their curricula. Students may also respond by increasing effort or altering study habits while in college. Any of these effects could lead the human capital outcomes for later cohorts to differ from those observed in the initial cohorts affected by the reform, even if the SES/exam score correlation, ρ^{exam}, and the degree of assortativity, ρ, are persistent over time.

Instability in the ability and quality distributions is potentially important in my setting given that the exam reform leads to larger changes in student/college matches than in related work. For example, regression discontinuity designs in the college selectivity literature capture the effects of school quality that correspond to a marginal reduction in a college's admission threshold. A marginal change in the threshold may have important implications for the student who gains admission, but it does not significantly alter the nature of the college's student body. However, a reform that leads to large-scale changes in student/college matches may compel schools to alter their curricula or support services in ways that marginal admission policy reforms do not.

Testing for the persistence of effects on student outcomes thus reveals something about the mechanisms that underlie students' human capital accumulation while in college. In particular, it reveals whether the complementarity between students and their colleges is fixed—as would arise if students' outcomes depend purely on their class rank or on a constant college value added—or whether student and college responses can change this complementarity.

In the final section of this paper, I examine the persistence these effects by measuring outcomes beyond the first few cohorts exposed to the redesigned admission exam. In particular, I examine persistence of the three stages of my analysis: 1) the strength of exam score/SES correlation, ρ^{exam}, as captured by equation (6); 2) the degree of assortativity, ρ, as captured by equation (7); and 3) the human capital effects as discussed in Propositions 1 and 2. I compare each of these outcomes for the first two post-reform cohorts to those observed in cohorts up to seven years after the overhaul.

3. Admission exam reform in Colombia

This section provides background on the Colombian college system, my related data sources, and a reform of the national admission exam. I show that the reform led to a sharp reduction in the correlation of exam scores with SES. This provides an instrument for student/college assortativity, which I examine in Section 4.

3.1. **Institutional background and data.** In this paper I use three administrative datasets that correspond to different parts of the Colombian higher education system.

The first dataset includes records from a national standardized exam called the ICFES, which Colombian students are required to take to apply to college.[17] The Colombian exam is generally analogous to the SAT in the U.S., but it is taken by nearly all high school graduates regardless of whether they plan to attend college. In this paper I use individual administrative records from the testing agency that cover 1998–2006 exam takers, which provide each student's test scores, high school, and background characteristics.

The second dataset includes enrollment and graduation records from the Ministry of Education that cover the near universe of colleges in Colombia. Colombia's higher education system consists of public and private institutions with varying selectivity and degree offerings. As in the U.S., admissions are decentralized; students apply to individual colleges and each institution controls its own selection criteria.[18] Nonetheless, the Ministry of Education tracks individual-level enrollment and graduation at nearly all colleges in the country. I use the Ministry's records on all enrollees from 1998–2012, which contain each student's institution, program of study, and dates of entry and exit.

[17] ICFES stands for Institute for the Promotion of Higher Education, the former acronym for the agency that administers the exam. The ICFES exam is now named Saber 11°.

[18] Unlike in the U.S., Colombian students apply not just to individual colleges but to college/major combinations. Below I discuss whether my effects are driven by college effects, major effects, or both.

My last data source is from the Ministry of Social Protection, which provides post-college earnings for all college enrollees. These data contain monthly earnings for any college enrollee employed in the formal sector in 2008–2012. From these records I calculate average daily earnings by dividing monthly earnings by the number of formal employment days in each month and averaging across the year.

I link the admission exam, college, and earnings records using individuals' names, birthdates, and ID numbers. The resulting dataset contains college enrollment and graduation outcomes for 1998–2006 exam takers and 2008–2012 formal sector earnings for college enrollees.[19]

3.2. **The 2000 admission exam reform.** The Colombian national exam was first administered in 1968 with the aim of supporting college admissions. As the exam achieved widespread coverage in the 1980s, the government additionally began to use admission exam results to evaluate high schools.

In the mid 1990s, policymakers concluded that the exam was poorly designed for the dual objective of college admissions and high school accountability. There was a perception that the exam rewarded innate ability and rote memorization more than the capacity to apply material one has learned. Critics thus argued that it was a poor measure of high school value added, and that scores were biased in favor of high SES students.

To address these concerns the testing agency completely redesigned the admission exam, altering the type of questions, tested subjects, and scoring system. The goal was to develop an exam that tested "competencies" rather than "content," and in doing so to align the exam with the high school curriculum and the skills that predict college success. Appendix A.3 gives further details on the changes to the exam structure, and it provides sample questions from the pre- and post-reform tests.

The new exam debuted in 2000 after more than five years of psychometric research. The exam overhaul was widely publicized in the preceding months including substantial coverage

[19] Appendix A.2 provides details on the coverage of each dataset and the merge process.

in *El Tiempo*, the leading Colombia newspaper. In its objective and publicity, the redesign of the exam was thus similar to the overhaul of the U.S. SAT that debuted in 2016.

3.3. **Reduction in the correlation of exam scores with SES.** This section shows that the 2000 overhaul led to a large and immediate decrease in the correlation of exam scores with individuals' SES. This corroborates assumption (6) in Section 2.

To do this I define an index X_i based on students' socioeconomic traits, where i denotes individuals. The characteristics I use are indicators for a mother with a secondary education, a student's high school, and a student's gender. The index X_i is the average pre-reform (1998–1999) admission exam score in cells defined by these characteristics.[20] There are over 7,500 high schools in my sample, so this index gives a finely grained prediction of a student's expected exam performance given her socioeconomic characteristics. On average, men, students with secondary-educated mothers, and students from top high schools have higher values of X_i.

Figure 2 shows how the socioeconomic index relates to scores on the mathematics subject of the admission exam for two cohorts before and after the reform.[21] The horizontal axis depicts the index X_i, and the vertical axis contains the math score. Each variable is normalized to mean zero and standard deviation one; this scaling is convenient because the linear regression coefficient—denoted by ρ^{exam} on the graph—is the correlation coefficient between the two variables.[22]

The dark solid lines show the non-parametric relationship between math scores and the socioeconomic index for the two pre-reform cohorts. The correlation between these variables is 0.55 among 1998 exam takers, and is nearly identical in 1999 at 0.56. This implies that students who are two standard deviations above the mean SES have average math scores roughly one standard deviation above the mean.

[20] This index is the average score across all exam subjects. I use a leave-one-out mean for each 1998–1999 exam taker so that X_i is not mechanically correlated with the outcomes examined below.

[21] Appendix A.2 provides details on the sample for Figure 2 and the remaining analyses in this paper.

[22] Equality of the regression and correlation coefficients is not strictly true for any individual cohort since I hold X_i constant over time, but the differences are minor.

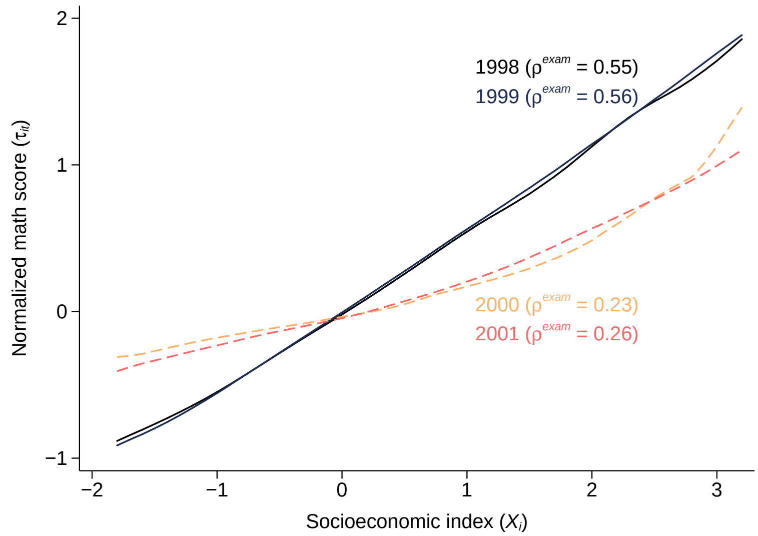


FIGURE 2. Math scores by admission exam cohort

Notes: This figure shows the non-parametric relationship between the socioeconomic index, X_i, and normalized math score, τ_{it}, for all exam takers in each cohort from 1998–2001. Both X_i and τ_{it} are normalized to mean zero and standard deviation one. It also displays the coefficient, ρ^{exam}, from the linear regression of τ_{it} on X_i for each cohort. See Appendix A.2 for details on the sample and variable definitions.

The dashed lines show that this relationship declines sharply in the two cohorts immediately following the reform; the correlation between the socioeconomic index and math scores falls to 0.23 in 2000 and 0.26 in 2001. Thus the exam reform led to substantially higher math scores for low SES students and relatively lower scores for high SES students. Put another way, the overhaul dramatically reduced ρ^{exam}, the correlation of math scores with X_i, consistent with assumption (6).

Figure 3 summarizes the reform's impact on other subjects of the exam. The height of each bar is analogous to ρ^{exam} in Figure 2; it is the correlation between the socioeconomic index, X_i, and the normalized subject score listed on the horizontal axis. The reform reduced this correlation by more than 50 percent in both math and physics, with more modest declines in other subjects. Yet for all subjects, the reform led to a sharp and meaningful decrease in the correlation of exam scores with SES.[23]

[23] Appendix Table A6 shows that SES/exam score correlation also declines when I consider the average exam score across all subjects. This finding also holds for different definitions of the socioeconomic index, X_i, including one based only on mother's education and one defined using only the 1998 exam cohort (instead of the 1998–1999 cohorts), which suggests that the patterns in Figure 3 are not due to mean reversion.

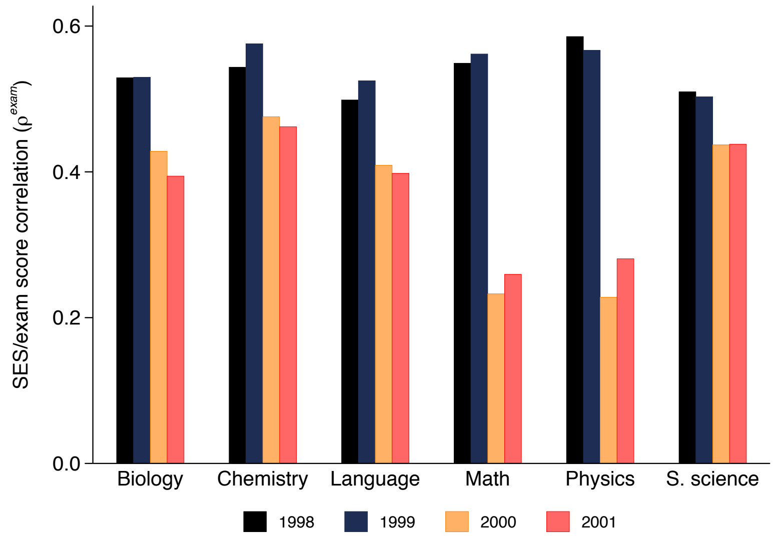


FIGURE 3. SES/exam score correlation by subject

Notes: The height of each bar is the coefficient, ρ^{exam}, from the linear regression of normalized math scores, τ_{it}, on the socioeconomic index, X_i, for each exam cohort and subject. See Section A.3 for details on the subjects included in the pre- and post-reform exams.

In sum, low SES students scored relatively higher on the post-reform exam, and high SES students scored correspondingly lower. Although this result suggests that the testing agency achieved its goal of designing an admission exam with less socioeconomic bias, it raises an additional question: what characteristics did the new exam measure? A second objective of the reform was to design a better measure of the competencies that predict success in college, but a decrease in the correlation of scores with SES could arise if the new exam simply contained more measurement error. Appendix A.4 explores the characteristics measured by the pre- and post-reform exams by taking advantage of the fact that a small number of students took both tests. It shows that the post-reform exam is a better predictor of these students' college *outcomes* once one controls for socioeconomic characteristics.[24] Thus the evidence is consistent with both of the major stated objectives of the overhaul: the redesigned exam led to a dramatic reduction in the correlation of admission scores and SES without substantially reducing its predictive power for college success.

[24] However, the differences in the predictive power of the pre- and post-reform exams are not statistically significant.

4. Assortativity results

This section shows that the exam reform described in Section 3 caused a decrease in student/college assortative matching in some regions of Colombia. I first define a set of "treated" regions in which the reform was ex ante more likely to affect allocation of students to colleges because the national exam plays a larger role in admissions. Then, I show that the reform led to a reduction in assortativity in treated regions relative to other regions, with low SES students displacing high SES students at top colleges. Finally, I discuss other potential confounding effects of the exam reform and provide evidence that the primary effect was a reduction in assortativity. This sets up a natural experiment that allows me to explore how a decrease in region-wide assortative matching affects human capital accumulation in Sections 5 and 6.

4.1. **Definition of treated and control regions.** I identify the exam reform's effects on assortativity by exploiting two features of the Colombian college system that create regional variation in the stakes of the national exam.

First, most of Colombia's large administrative regions have a flagship public university that is a major player in the local market.[25] College attendance is highly regional, and in some cases the flagship school enrolls more than one-third of all area college students. Flagships are substantially less expensive than comparable private universities and give large tuition discounts to low-income students. Since financial aid markets are underdeveloped, the local flagship is often the only top college available to low SES students. As a result, flagship universities are the most selective colleges in the country. For example, the flagship university in Bogotá, Universidad Nacional de Colombia, typically admits less than ten percent of applicants, while the top private college, Universidad de Los Andes, has an admission rate near 50 percent.

The second relevant feature of the Colombia college system is that flagship universities control their own admission criteria. The majority of flagships admit students solely based

[25] Colombia has 33 administrative departments (*departamentos*) that I call regions.

on the national admission exam scores. These colleges compute major-specific weighted averages of the national exam subject scores and admit the highest ranking students up to a predetermined quota. Other flagships, however, require applicants to take the university's own exam, and some also consider high school grades or personal interviews. For example, Universidad Nacional de Colombia administers its own entrance exam, while Universidad del Valle, the flagship school in Cali, uses only national exam scores for admission.

I classify Colombia's 33 administrative regions as treated or control depending on the pre-reform admission method of their flagship university. For this I collected data on the pre-reform admission methods at each flagship university from historical websites and student regulations.[26] Black dots in Figure 4 represent cities with a flagship university that uses only national exam scores for admissions. White dots are cities in which the flagship uses its own entrance exam or other admissions criteria. Treated regions have flagships with national exam admissions and are shaded red in Figure 4. The light-colored control regions are those with other flagship admission methods. I define treatment for the sparsely-populated southeastern regions with no universities using the closest flagship to their capital city.[27]

Table 1 displays summary statistics for pre-reform college enrollees in the 22 treated and 11 control regions. Both areas have about 90,000 enrollees per year, but control regions have more colleges per student. Students in treated regions are more likely to attend public colleges, while control students are more likely to enroll in technical schools or institutes rather than universities.

The bottom of Table 1 shows how the features of the Colombian college system described above are useful for identification. Most students attend college in their own region, including 57 percent of treated students and 84 percent of control students. Flagship universities in both areas enroll a large fraction of all students, but there is substantial variation in exposure to flagships that use the national exam. Nearly one-third of treated students attend national exam admission flagships, compared with just two percent of control students.

[26] Appendix A.5 provides details on the flagship universities, data sources, and admission methods.

[27] Bogotá is its own administrative region. Figure 4 does not display the island region of San Andrés y Providencia, which I define as a control region.

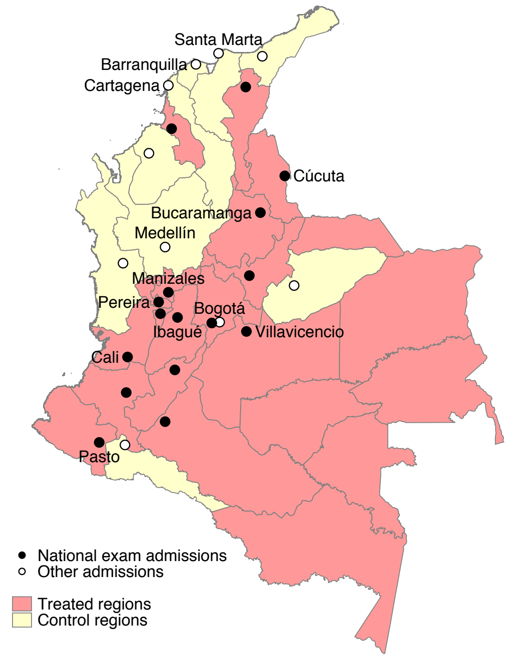


Figure 4. Definition of treated and control regions in Colombia

Notes: Dots represent flagship universities. Treated regions (in red) are those with flagships that use only the national exam scores for admissions (black dots). Control regions (in light yellow) are those with flagships that use other admission methods (white dots). Bogotá, which is its own administrative region, is a control region. The map does not show the island region of San Andrés y Providencia, which I define as a control region. I define treatment for regions with no universities using the closest flagship to their capital city. See Appendix A.5 for details.

Table 2. Summary statistics for pre-reform (1998–1999) college enrollees

	Treated regions	Control regions
Regions	22	11
Colleges in region	111	185
College enrollees/year	89,143	89,098
Enrolled in a public college	0.51	0.37
Enrolled in technical school/institute	0.21	0.36
Enrolled in region	0.57	0.84
Enrolled in any flagship university	0.33	0.16
Enrolled in national exam flagship	0.31	0.02

Notes: See Figure 4 and Appendix A.5 for the definitions of flagship universities and treated/control regions. The number of college enrollees per year is the average number calculated from the 1998–1999 cohorts.

The combination of regional markets and decentralized flagship admissions creates geographic variation in the potential impact of the national exam reform. The exam redesign was *ex ante* more likely to affect college admissions in treated regions, where flagship universities with national exam admissions comprise a large fraction of the market. The new exam altered the ranking of students in flagship admission pools and thus had the potential to alter the type of students received by many colleges in the region. The new ranking mattered less at flagship universities that used their own exams, so the reform was less likely to affect college admissions throughout control region markets.

4.2. **Effects on assortativity.** I use a differences in differences specification to show that the exam reform reduced market-wide assortative matching in treated regions relative to other areas of the country. My specification differs from a standard differences in differences analysis in that it measures changes in a slope coefficient rather than changes in a level. Specifically, the goal is to measure the reform's effects on the correlation of SES with college quality, represented by the coefficient ρ from equation (4) in Section 2. This specification follows Card and Krueger (1996), who ask how state-level policies affect another slope—the return to education.

To provide intuition for this specification, it is useful to consider a two-step regression procedure. The first step is modified version of the equation that defines assortativity:

$$W_c = \gamma_{rt} + \rho_{rt} X_i + u_{icrt}, \tag{16}$$

W_c is the observable quality of college c, defined as its average admission exam score using the two pre-reform cohorts in my data (1998–1999). X_i is the socioeconomic index defined in Section 3.3, which is individual i's expected pre-reform admission exam score given her background characteristics. γ_{rt} are fixed effects for cells defined by region r and admission exam cohort t. I define region as the administrative department in which a student attended high school, and I include the 1998–2001 exam cohorts. u_{icrt} is the residual. This equation is analogous to equation (4) except it calculates a different value of assortativity, ρ_{rt}, for each region and exam cohort.

The second step compares changes in the assortativity parameters across cohorts and across treated and control regions:

$$\rho_{rt} = \gamma_r + \gamma_t + \beta^{\rho} \delta_{rt} + e_{rt}. \tag{17}$$

The regression includes region dummies, γ_r, cohort dummies, γ_t, and a residual, e_{rt}. The coefficient of interest, β^{ρ}, is on the variable δ_{rt}, which is an indicator equal to one for graduates from high schools in treated regions and post-reform cohorts (2000–2001). δ_{rt} equals zero for all pre-reform cohorts (1998–1999) and for all graduates from control regions. The coefficient β^{ρ} measures how the change in assortativity, ρ, between the pre- and post-reform cohorts in treated region compares to that in control regions.

I estimate a single-step specification that is derived by plugging equation (17) into (16):

$$W_c = \gamma_{rt} + (\gamma_r + \gamma_t + \beta^{\rho}\delta_{rt})X_i + e_{icrt}. \tag{18}$$

Equation (18) is a differences in differences regression on the slope ρ rather than on a level. Since both W_c and X_i are normalized to mean zero and standard deviation one, the coefficient β^{ρ} captures the differential change in the correlation of SES and college quality in treated and control regions. I cluster standard errors in all regressions at the region level.[28]

Since the exam overhaul reduced the correlation of scores with observable characteristics, X_i, the main prediction is $\beta^{\rho} < 0$; the reform should reduce the correlation of X_i with college quality, W_c, in treated regions relative to control regions. This corresponds to the "first-stage" assumption (7) in Section 2.

Figure 5 plots coefficients from an event-study version of equation (18), which calculates separate coefficients β^{ρ}_t for each exam cohort t. The horizontal axis shows the four exam cohorts, and the vertical axis displays the β^{ρ}_t coefficients. The regression omits the interaction between the treatment variable δ_{rt} and the dummy for 1999 graduates; this sets $\beta^{\rho}_{1999} = 0$ and makes all other coefficients relative to the 1999 cohort. The identification assumption is parallel trends in treated and control regions in the counterfactual outcomes absent the reform. This implies that the pre-reform coefficients (1998–1999) should be similar, while sharp changes in outcomes for the post-reform cohorts (2000–2001) support a causal interpretation of my results.

[28] Appendix Table A9 shows that my main results are robust to the wild t bootstrap procedure recommended by Cameron, Gelbach and Miller (2008) for contexts with a relatively small number of cluster.

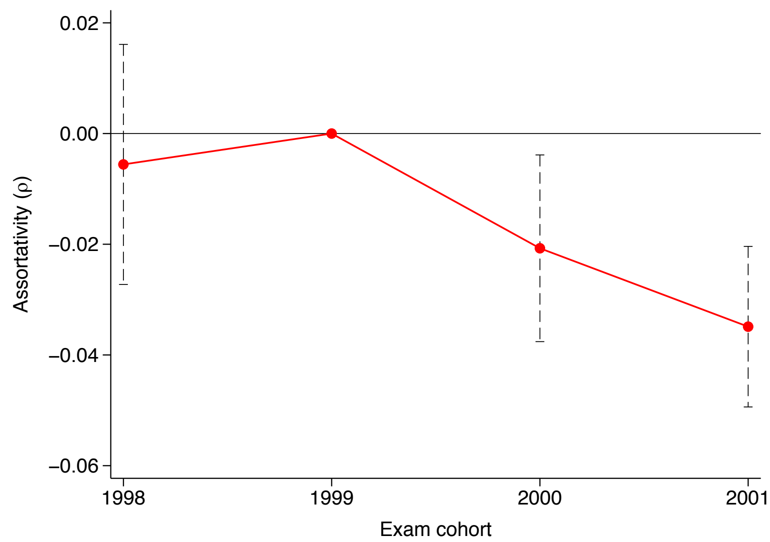


FIGURE 5. Exam reform effects on assortativity

Notes: This figure plots event study coefficients β_t^ρ from equation (18), with the interaction between the treatment variable δ_{rt} and the dummy for 1999 graduates as the omitted group. Dashed lines are 90% confidence intervals using standard errors clustered at the region level.

There is little evidence of differential pre-trends in assortativity in the two cohorts before the reform. The average difference in ρ between treated and control regions is only 0.005 smaller in 1998 than it is in 1999, and this difference is not statistically significant. In 2000, the first year of the new exam, the correlation between student and college traits falls sharply in treated regions relative to control regions. By the 2001 cohort, the gap between treated and control correlations is more than 0.03 points lower it was in 1999. In other words, assortative matching of students and colleges—as defined by pre-reform characteristics—fell in treated regions relative to control regions.

Table 3 presents regression estimates analogous to Figure 5. Column (A) depicts the benchmark estimate of β^ρ from equation (18). The point estimate suggests that the exam reform led to a 0.025 point decrease in the correlation of SES and college quality in treated regions relative to control regions.

The remaining columns of Table 3 report results using different definitions of assortativity. Column (B) is identical to column (A), except it defines the dependent variable, W_c, as the average pre-reform admission exam score at the college-major level rather than the college level. This reflects the fact that in Colombia admissions are to college-major pairs. Thus

Table 3. Exam reform effects on assortativity

Dependent variable: College quality (W_c)

Treated regions and cohorts (δ_{rt}) $\times$ ...	(A) Benchmark specification	(B) Definition of observable characteristics: W_c defined at college-major level	(C) Definition of observable characteristics: X_i defined by mother edu. only	(D) Definition of treated & control regions: Closest flagship to municip.	(E) Definition of treated & control regions: Pre-reform enrollment in municip.
Socioeconomic index (X_i)	-0.025^{***}	-0.020^{**}	-0.014^{**}	-0.028^{***}	-0.031^{***}
	(0.007)	(0.009)	(0.006)	(0.007)	(0.008)
N	714,071	714,071	714,071	714,071	714,071
R^2	0.256	0.279	0.118	0.270	0.270
# regions	33	33	33	33	33
Mean assortativity	0.429	0.496	0.171	0.437	0.437

Notes: The table reports estimates of β^ρ from equation (18). Regressions in columns (D) and (E) replace region fixed effects with municipality fixed effects. See the text for descriptions of the variables in each column. Parentheses contain standard errors clustered at the region level. Mean assortativity is calculated from the 1998–1999 cohorts.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

it is possible that the exam reform led to changes in the distribution of students' majors in addition to changes in the distribution of their colleges. Columns (A) and (B) show that the effects of the reform on assortativity are similar using the college and college-major definitions of quality. Further, Appendix Table A7 shows that when I define W_c at the major level only, the analogous coefficient is not statistically different from zero. This suggests that the primary effect of the exam reform was a change in the schools students attended rather than a change in their majors.[29]

Column (C) presents results using a different definition of the socioeconomic index, X_i. The benchmark regression in column (A) defines X_i as the mean pre-reform admission score in cells defined by mother's education, gender, and indicators for a student's high school. This gives a finely grained measure of an individual's socioeconomic status. In column (C),

[29] This result contrasts with the finding in Kirkebøen, Leuven and Mogstad (2016) that the earning returns to college-major pairs are primarily driven by major effects. One potential explanation for this result is a difference between the higher education markets in Norway—the setting for the Kirkebøen, Leuven and Mogstad (2016) paper—and Colombia. Anecdotally, the Colombian higher education system features greater variation in college reputation, while colleges in Norway differ less in their perceived rankings. The results in this paper also align with those in MacLeod et al. (2015), who find similar effects of the introduction of a Colombian college exit exam using college and college-major definitions of school reputation.

I define X_i simply as a dummy for whether a student's mother has a secondary education.[30] The effects on assortativity are similar to those in the benchmark specification. This shows that the reform led to changes in distribution of college quality using even a rudimentary definition of SES; however, the more finely grained measure of X_i is useful for analyzing the distribution of outcomes below.

Columns (D) and (E) show that the effects on assortativity are robust to different definitions of treated and control regions. The benchmark specification defines the treatment variable, δ_{rt}, by whether or not the flagship university in the region where a student attended high school uses only the national exam score for admissions. Column (D) defines treatment instead using the closest flagship university to a students' municipality. This reflects the fact some students are likely to attend college out-of-region if their high school is close to a major city in a different region. The first stage effects on assortativity are similar using this more granular definition of treatment.[31] Column (E) also uses a municipality-level definition of δ_{rt}, but it is even more granular in that it allows for different intensities of treatment. Treatment is defined using the fraction of pre-reform college students from each municipality who enrolled in a flagship university using national exam admissions.[32] The effects of the reform on assortativity are slightly stronger using this intensity of treatment variable. Nonetheless, the similarity of the coefficients in columns (A) and (E) suggests that the benchmark specification captures the first order effects of the exam reform.

4.3. **Alternative hypotheses.** This section considers other effects of the exam overhaul, and it argues that these alternative hypotheses are less likely to impact students' human capital accumulation than the reduction in assortativity.

---

[30] I normalize this dummy to standard deviation one to make the coefficient in column (C) comparable to those in others columns.

[31] The regressions in columns (D) and (E) of Table 3 replace the region-level fixed effects in specification (18) with municipality-level fixed effects.

[32] Specifically, in column (E) δ_{rt} equals zero for all pre-reform cohorts, and for post-reform cohorts it equals the fraction of college enrollees from the 1998–1999 cohorts in municipality r who enrolled in a flagship that used only national exam scores for admissions. I normalize δ_{rt} to be mean zero and standard deviation 0.5, which is approximately equal to the standard deviation of δ_{rt} in the benchmark specification. This makes the coefficients in columns (A) and (E) comparable in magnitude.

The key assumption for Proposition 1 is that the distributions of academic ability, A_i, and college quality, Q_c, are stable across exam cohorts. Instability of the academic ability distribution may arise if the reform affected the probability that students enrolled in any college, or if it induced them to retake or study harder for the exam. Instability of the college quality distribution could arise if the number of students enrolling in each college changed with the reform, which could stem from changes in school quotas or if students selected colleges outside of their region.

To explore these alternative hypotheses, I use a standard differences in differences specification. This specification is identical to the second step regression (17), but it considers other dependent variables, y_{irt}:

$$y_{irt} = \gamma_r + \gamma_t + \beta\delta_{rt} + e_{irt}. \tag{19}$$

As above, this regression includes both region r and cohort t dummies, and the treatment variable, δ_{rt}, is equal to one for graduates from high schools in treated regions and post-reform cohorts. The coefficient of interest, β, measures the differential change in the outcome y_{irt} between treated and control region. I also estimate a specification that fully interacts the right-hand side of equation (19) with dummies for quartiles q of the advantaged index, X_i. This yields separate coefficients β_q for students from different quartiles of the socioeconomic distribution.[33]

Table 6 presents estimates of β and β_q for dependent variables that correspond to other potential effects of the exam reform. Each column of the table reports coefficients from two different regressions. The top row shows estimates of β from the standard differences in differences regression (19). The bottom four rows present estimates of β_q from the version with fully-interacted dummies for quartiles q of the socioeconomic index.

Column (A) in Table 6 shows that the primary effect of the reform was on *where* students went to college, not *whether* they went to college. This regression expands the sample to

[33] The coefficients β_q are identical to those that would be obtained by estimating equation (19) separately for each quartile q.

Table 4. Alternative hypotheses

	(A)	(B)	(C)	(D)	(E)
	Dependent variable				
Treated regions and cohorts (δ_{rt}) $\times$ ...	Enrolled in any college	Graduating from HS this cohort	Enrolled within one year	Kilometers between HS & college	Enrolled in different region
All students	-0.004	0.001	-0.001	-3.815	-0.014^{*}
	(0.008)	(0.009)	(0.010)	(3.082)	(0.008)
X_i top quartile	-0.019	-0.007	-0.005	-1.294	0.003
	(0.013)	(0.015)	(0.007)	(3.452)	(0.009)
X_i quartile 3	-0.003	0.009	0.012	-7.178^{***}	-0.026^{***}
	(0.009)	(0.007)	(0.010)	(2.537)	(0.008)
X_i quartile 2	0.008	0.010	0.004	-8.222^{**}	-0.034^{***}
	(0.008)	(0.008)	(0.015)	(3.313)	(0.008)
X_i bottom quartile	-0.000	-0.008	0.013	-2.037	-0.026^{*}
	(0.007)	(0.012)	(0.023)	(7.517)	(0.013)
N	1,878,799	714,071	714,071	714,071	714,071
# regions	33	33	33	33	33
Dependent var. mean	0.358	0.790	0.426	82.734	0.298

Notes: The top row reports estimates of β from equation (19). The bottom four rows report β_q coefficients from a version of (19) that is fully interacted with dummies for quartiles q of the socioeconomic index, X_i. The column header shows the dependent variable for each regression. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 1998–1999 cohorts.

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

include all exam takers—not just those who attended college—and the dependent variable is an indicator equal to one if a student attended any college in my records. The estimated effect on any college enrollment is small and insignificant for the full sample, and there are no statistically significant effects on the probability of college enrollment in any quartile of the socioeconomic index, X_i. One explanation for this is that my instrument primarily affects admission to selective schools, and Colombian college markets have a large, open-enrollment sector where students can enroll if they are not admitted to top colleges. It is also parallels the findings in research on other large-scale admission policies that primarily affect admission to selective colleges. For example, other work has found that affirmative action bans (Hinrichs, 2012) and percent plan admission rules (Daugherty, Martorell and McFarlin, 2014) have little effect on the extensive margin of college enrollment.

Columns (B) and (C) show that the reform had little effect on students' likelihood of repeating the exam or delaying their enrollment into college. The dependent variable in column (B) is an indicator equal to one for students who took the exam in their year of high school graduation. The dependent variable in column (C) is a dummy equal to one for students who enrolled in college within one year of taking the admission exam. There are no statistically significant effects on the average or distribution of these variables, which suggests that the reform did not substantial alter the probability that students retook the exam or delayed their college enrollment.

By contrast, columns (D) and (E) present evidence that the reform had some effects on student mobility. Column (D) shows no average effect on the distance between students' high schools and colleges, but there was a reduction in this distance for students in the middle of the socioeconomic distribution. This pattern is similar when I use a dependent variable equal to one for students who attended college in a region different from that of their high school (column (E)), although in this case the full-population coefficient is also marginally significant. In both cases, there is evidence the reform may have induced some students to attend college closer to home.

The results in columns (D) and (E) raise the concern that there may have differential changes in average college quality across treated and control regions. To explore this further, Table 5 shows the effect of the reform on four different measures of college quality: my primary measure of quality, W_c, which is the average pre-reform admission score at each college (column (A)); an indicator equal to one for colleges in the top ten percent of W_c (column (B)); an indicator for whether an institution offers university-level training as opposed to technical training (column (C)); and an indicator for colleges that are selective, meaning that they have more applicants than slots (column (D)).[34] The top row shows that for the full population of college enrollees, there were no statistically significant changes in average college quality across all dependent variables. This suggests that even if the reform affected

[34] Specifically, selective colleges are those for whom the number of applicants exceeded the number of offered slots in a 2002 report from the testing agency entitled *Estadísticas de la Educación Superior*. I could not find information on the number of slots in a pre-reform year.

TABLE 5. Distribution of college quality

	(A)	(B)	(C)	(D)
	Dependent variable			
Treated regions and cohorts (δ_{rt}) $\times$ ...	College quality (W_c)	College in the top W_c decile	University-level institution	Selective college
All students	0.006	−0.005	−0.009	0.022
	(0.021)	(0.003)	(0.007)	(0.017)
X_i top quartile	−0.023	-0.010^{*}	-0.019^{***}	0.003
	(0.024)	(0.006)	(0.006)	(0.019)
X_i quartile 3	0.032	0.005	−0.001	0.027
	(0.023)	(0.003)	(0.009)	(0.018)
X_i quartile 2	0.046^{*}	−0.000	0.007	0.044^{***}
	(0.024)	(0.002)	(0.009)	(0.014)
X_i bottom quartile	0.050^{***}	−0.001	−0.001	0.041^{**}
	(0.017)	(0.002)	(0.009)	(0.017)
N	714,071	714,071	714,071	714,071
# regions	33	33	33	33
Dependent var. mean	−0.000	0.101	0.715	0.476

Notes: The top row reports estimates of β from equation (19). The bottom four rows report β_q coefficients from a version of (19) that is fully interacted with dummies for quartiles q of the socioeconomic index, X_i. The column header shows the dependent variable for each regression. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 1998–1999 cohorts.

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

student mobility, it did not have a substantial effect on the average change in college quality across regions.

By contrast, the bottom four rows of Table 5 show that the reform affected the allocation of college quality across students. Although not all effects are statistically significant, there is a consistent pattern that the reform decreased the quality of colleges attended by the highest SES students in treated regions. Correspondingly, average college quality increased for treated region students with the lowest values of the socioeconomic index. This result shows that the changes in assortativity documented in Figure 5 and Table 3 hold for other measures of college quality.

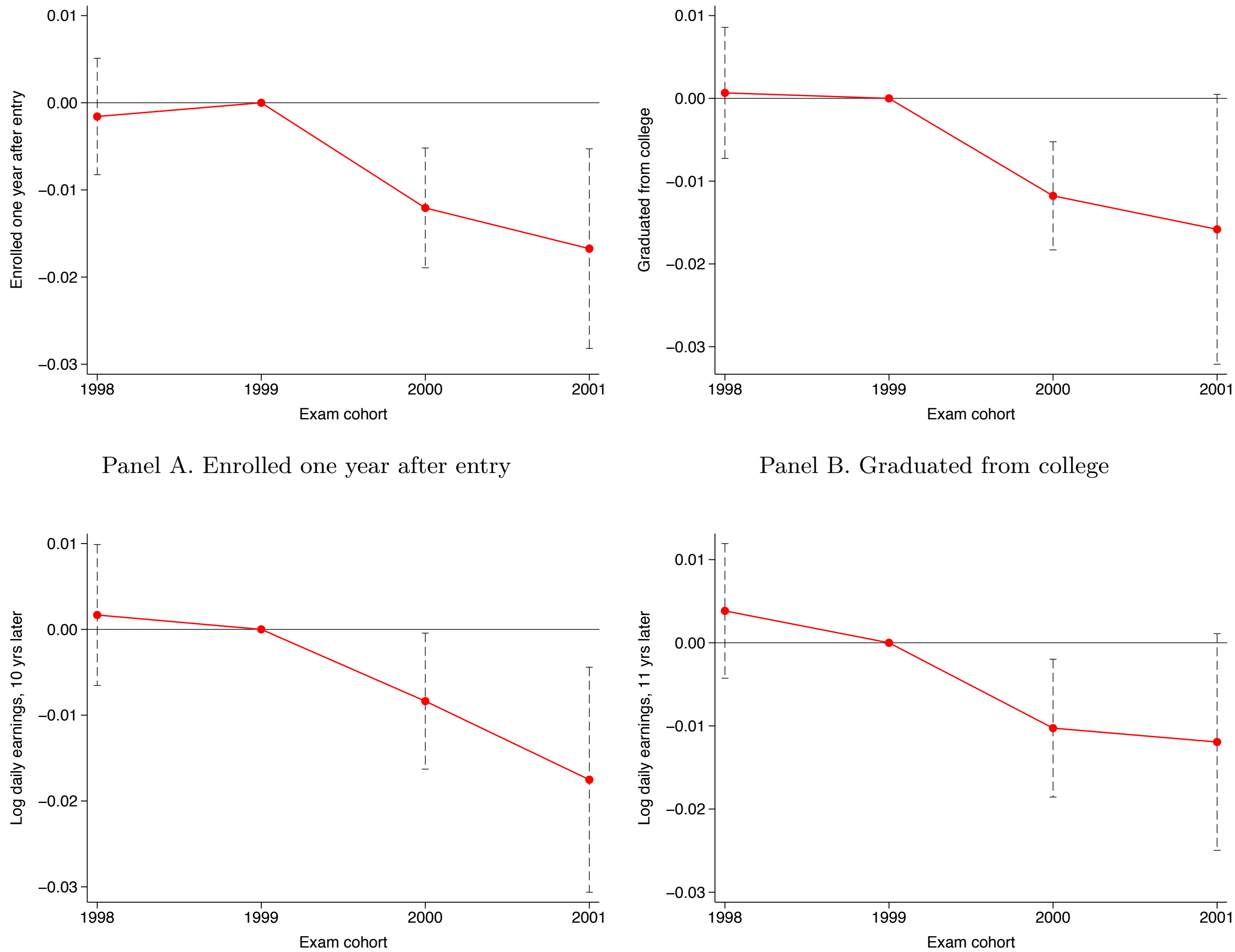
In sum, the results from Tables 6 and 5 suggest that the exam reform did not have substantial effects on the distribution of academic ability, and though there is some evidence of student mobility responses, these effects did not lead to average changes in college quality across regions. Further, the strongest effects on mobility occur in the middle of the SES distribution, whereas the strongest effects on the allocation of college quality arise for the highest and lowest values of X_i. Taken together, these results suggest that the first order effect of the exam reform—at least for the highest and lowest SES students—was a reallocation of college quality, rather than a change in its aggregate value.

5. Human capital results

This section shows that the reduction in assortativity from the Colombian exam reform led to a region-wide reduction in graduation rates and post-college earnings. Further, I show that these negative effects are driven by both the highest SES and the lowest SES students, with the latter result providing evidence of college mismatch. Finally, I discuss implications for Propositions 1 and 2 from the framework: the results reveal a positive complementarity between SES and college quality, and they provide mixed support for the validity of the "conditional on observables" assumption in the mismatch literature.

5.1. **Graduation and earnings effects of the reform.** I use the standard differences in differences specification (19) to measure the effect of the reform on different measures of human capital accumulation, Y_{irt}. The β coefficient from this regression corresponds to the *average reallocation effect* from equation (8) in Section 2. This measures the average market-wide change in human capital in treated regions relative to control regions.

Figure 6 displays event-study coefficients β_t for several measures of human capital. These graphs are analogous to Figure 5, but they come from event-study versions of equation (19) rather than equation (18). Panels A and B use measures of college completion as outcome variables: persistence and graduation. Panel A shows that the probability of being enrolled in college one year after entry changes little across regions between the 1998 and 1999 cohorts, but it falls sharply in the first year of the exam reform. Persistence rates fell by about 1.5 percentage points among the 2000–2001 cohorts in treated regions relative to control students. Panel B shows a similar pattern and magnitude for graduation rates. These panels suggest that the reallocation of college quality caused some students in treated



Panel A. Enrolled one year after entry

Panel B. Graduated from college

Panel C. Log daily earnings, 10 years later

Panel D. Log daily earnings, 11 years later

FIGURE 6. Persistence, graduation, and earnings

Notes: This figure plots event study coefficients β_t from equation (19), with the interaction between the treatment variable δ_{rt} and the dummy for 1999 graduates as the omitted group. Dashed lines are 90% confidence intervals using standard errors clustered at the region level.

regions to drop out of college within the first year, and this effect remained stable through college graduation.

Panels C and D present analogous results using log daily earnings as the dependent variable. Panel C uses earnings measured ten years after students took the admission exam, and Panel D shows earnings eleven years after the exam.[35] The earnings results mirror those for college completion. There are no statistically different changes between treated and control

[35] I observe earnings in 2008–2012, so ten and eleven years post-exam are the two years for which I observe earnings for all four cohorts. Appendix Table A8 shows that the reform had little effect on the probability of being employed in the formal sector.

Table 6. Persistence, graduation, and earnings

	(A)	(B)	(C)	(D)
	Dependent variable			
Treated regions and cohorts (δ_{rt}) $\times$ ...	Enrolled one year after entry	Graduated from college	Log daily earnings, 10 yrs later	Log daily earnings, 11 yrs later
All students	-0.014^{**}	-0.014^{*}	-0.014^{**}	-0.013^{**}
	(0.005)	(0.007)	(0.006)	(0.006)
X_i top quartile	-0.017^{***}	-0.022^{***}	-0.013	-0.016^{*}
	(0.005)	(0.007)	(0.009)	(0.008)
X_i quartile 3	-0.010	-0.004	-0.002	0.003
	(0.006)	(0.008)	(0.005)	(0.007)
X_i quartile 2	-0.006	-0.001	0.001	0.001
	(0.007)	(0.010)	(0.006)	(0.006)
X_i bottom quartile	-0.022^{**}	-0.022^{**}	-0.016^{*}	-0.013^{**}
	(0.008)	(0.009)	(0.008)	(0.006)
N	714,071	714,071	354,034	361,044
# regions	33	33	33	33
Dependent var. mean	0.649	0.394	10.269	10.374

Notes: The top row reports estimates of β from equation (19). The bottom four rows report β_q coefficients from a version of (19) that is fully interacted with dummies for quartiles q of the socioeconomic index, X_i. The column header shows the dependent variable for each regression. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 1998–1999 cohorts.

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

regions from 1998 to 1999, although there is a slightly larger pre-trend in Panel D. Earnings decline by about one percent in the 2000 treated cohort relative to control students, and this gap widens further in 2001. These results suggest that the negative effects of reallocation on students' college competition also affect their labor market outcomes.

Table 6 presents regression estimates for these four measures of human capital accumulation. This table has the same structure as Tables 6 and 5. The top row displays the estimate of β from equation (19), which gives the regression analog of the results in Figure 6. The magnitudes of the full-population treatment effects are similar across all outcome variables, with 1.4 percentage point declines in persistence and graduation rates, and 1.4 percent reductions in earnings.

The bottom four rows of Table 6 show estimates of β_q, which come from a version of equation (19) that is fully interacted with dummies for quartiles q of the socioeconomic index,

X_i. This specification yields reallocation effects for individuals with different socioeconomic backgrounds, and it corresponds to equation (14) in Section 2. The results suggest that the market-level declines in human capital accumulation are driven by students both at the top and bottom of the SES distribution. Persistence and graduation rates fall by more than two percentage points for treated students relative to control students in the top quartile of X_i. The decline in college completion is similar in magnitude for the lowest SES students in treated regions, although it is larger in percentage terms as these students are less likely to finish college. Both the top and bottom quartile students experience roughly 1.5 percent declines in earnings relative to control students.

5.2. **Implications for Proposition 1.** Proposition 1 states that a change in assortativity should have no effect on market-level human capital unless there is a complementarity between student and college characteristics. The results in Figure 6 and the top row of Table 6 provide clear evidence that the reallocation of student/college matches can affect market-wide human capital accumulation. Thus they are consistent with the existence of a complementarity between students and colleges.

The key assumption necessary to interpret these results as evidence of a complementarity is that the distributions of academic ability and college quality are not affected by the reform. Tables 6 and 5 provide support for this assumption, but the main outstanding consideration is that the reform altered peer composition at colleges. If peer effects are important for students' human capital accumulation, the graduation and earnings effects may be due to changes in peer composition rather than a student/college complementarity.[36]

There are two arguments that suggest peer effects are not completely driving the declines in human capital accumulation. First, peer effects must take a complicated functional form in order to explain the full pattern of my results. As shown in Table 1, strictly linear peer effects—in the spirit of Epple and Romano (1998*b*)—would lead to a zero reallocation effect, as any gains to some students are exactly offset by losses to other students. Furthermore,

[36] Peer effects may be broadly thought of as a consequence of the match between a student and her college, but it is not strictly a complementarity according to my definition in Section 2.

Table 6 shows that both low SES and high SES students experienced negative human capital effects even though the reform led to opposite changes in their peer quality.[37] Thus peer effects must be oppositely-signed for low SES and high SES students in order to explain the effects of the reform.

Second, Appendix Table A10 presents an additional analysis that argues against a model of pure peer effects. This analysis exploits the fact that some high SES students choose to attend lower-ranked schools even when they could be admitted to top colleges. I use pre-reform characteristics to divide high SES students into two types: those who were likely to choose top colleges, and those who were likely to choose lower-ranked colleges. I first show that the former group of students experience changes in both college quality and peer quality as a result of the reform, while the latter group experienced only changes in peer quality. I then show that the negative graduation and earnings effects are primarily driven by the students who saw changes in both college and peer quality. This suggests that, at least for high SES students, peer effects cannot fully explain my results, and it provides further support for the existence of a complementarity between student and colleges characteristics.

5.3. **Implications for Proposition 2.** Proposition 2 contains predictions about the sign of the student/college complementarity under a key assumption from work on college mismatch. The best existing evidence for mismatch assumes that students and colleges are randomly matched conditional on their observable characteristics, as discussed in Section 2.5. Under this assumption, Proposition 2 shows that the student/college complementarity is positive in the Colombian context, and thus the reduction in assortativity should lower market-wide graduation rates and earnings. However, this assumption also implies that there should also be no mismatch from the Colombian reform, as the observed pre-reform return to college is positive for all students.

[37] Average peer quality as defined by pre-reform exam scores increased for low SES students and decreased for high SES students. However, it is possible that these signs are flipped on other dimensions of peer characteristics.

To test this "conditional on observables" assumption, I compare the total effects of the reform, as given by Table 6, to those that would be predicted by distributional changes alone. I compute this *distribution effect* in two steps. First, I calculate the average pre-reform outcome in cells defined by quantiles of the socioeconomic and college quality indices, which I denote by $\hat{Y}_{ic}$:[38]

$$\hat{Y}_{ic} = E_0[Y_{ic}|X_i = x, W_c = w]. \tag{20}$$

Second, I estimate the β and β_q coefficients from standard differences in differences regressions that instead use $\hat{Y}_{ic}$ as the dependent variable. These estimated coefficients give the *distribution* effect of the reform described in Section 2.5, which is the effect from changes in the distribution of matches holding constant the observable pre-reform match outcomes.

Figure 7 plots the total and distribution effects for log earnings measured ten years after the admission exam. The black bars show the total effects, which are identical to the β and β_q coefficients from the log earnings regressions in column (C) of Table 6. The lighter red bars show distribution effects from the above procedure.

The top of Figure 7 shows that the total and distribution earnings effects are similar for the full population of students. The total effect of the reform was a 1.4 percent reduction treated region earnings relative to control regions. Distribution changes alone can explain a one percent reduction. This suggests that complementarity on *observed* student/college characteristics is, at least to a first-order, predictive of the aggregate reallocation effects. In other words, the fact that the observed college quality in Figure 1 increases with SES is not exclusively due to sorting; top colleges may be partially better prepared to educate high SES students.

However, the bottom of Figure 7 shows that distribution effects cannot fully explain the results for different quartiles of the socioeconomic index, X_i. While the total and distribution effects are similar for the top quartile of X_i, distributional changes alone cannot account for

[38] Specifically, I use 40 quantiles of the socioeconomic index, X_i, and ten deciles of observable college quality, W_c. This yields a 10×10 matrix of mean outcomes for each quartile of the socioeconomic index. In this notation, 0 stands for the pre-reform cohorts (1998–1999).

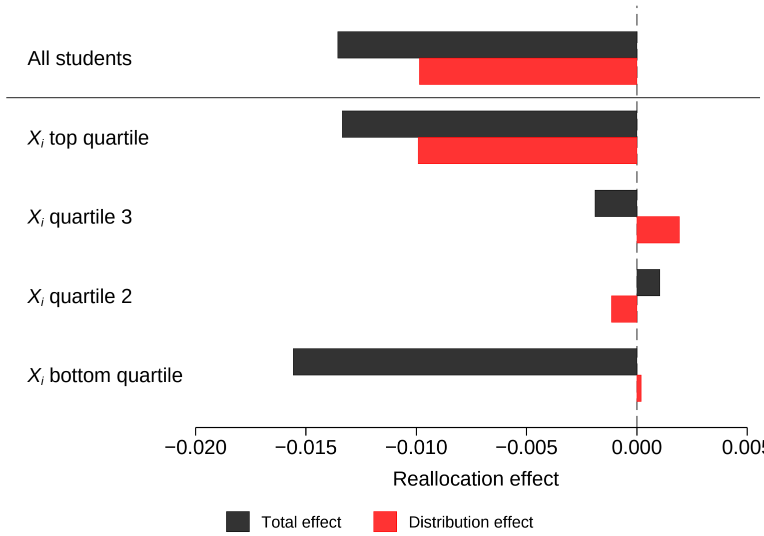


FIGURE 7. Total and distribution effects for log daily earnings, 10 years later

Notes: The black bars plot the point estimates from column (C) in Table 6. The light-red bars plot coefficients from identical regressions, except the dependent variable is $\hat{Y}_{ic}$ as defined by equation (20). In other words, it is the average log earnings within cells defined by 40 quantiles of the socioeconomic index, X_i, and ten deciles of observable college quality, W_c, calculate using only the 1998–1999 cohorts.

the mismatch effects observed for the lowest SES students. In particular, distribution effects predict a small increase in earnings for students in the lowest quartile of X_i, but the reform actually led to a 1.6 percent decline in earnings for these students.[39]

Appendix Table A11 shows that the patterns for log earnings in Figure 7 are similar for college persistence or graduation outcomes.[40] Thus the Colombian admission exam reform provides mixed support for the "conditional on observables" assumption employed in mismatch research; this assumption predicts the market-wide impact of the reform to a first-order level, but it fails to capture the full distribution of effects across SES groups.

There are two takeaways for the results related to Proposition 2. First, they suggest that *observed* differences in graduation rates or earnings for similar students at different colleges may be due in part to selection bias. Exogenous variation in student college matches is important to test the mismatch hypothesis, as Arcidiacono and Lovenheim (2016) note.[41]

[39] Figure 1 and Figure 7 correspond to slightly different distribution effects because I use a non-parametric calculation for the estimates in Figure 7.

[40] In fact, the distribution effects for persistence and graduation rates are slightly less predictive of the total effects for the full population of students.

[41] Rothstein and Yoon (2008) also make this point in the context of law school admissions.

My results provide evidence of mismatch effects without relying on the "conditional on observables" assumption common in other studies in this literature.

Second, it is worth exploring why the mismatch effects in my context may not arise in papers in the college selectivity literature. This research often finds positive earnings effects of admission to a selective college for low income students. One potential explanation for this apparent discrepancy is that the Colombian reform may have led some students to attend colleges at which they were well below the target teaching level. In selectivity studies, all admitted students—whether high or low SES—are qualified for admission by the college's own academic standards. The Colombian exam reform, by contrast, led to large changes in the distribution of students' academic preparation at selective college.[42] Thus the reform may have caused some students to attend colleges at which they would have been deemed unqualified prior to the reform.

This suggests that mismatch may only arise when students are substantially underprepared for a college's curriculum. Other contexts that exhibit mismatch have similarly large differences in students' incoming academic preparation (e.g, Arcidiacono, Aucejo and Hotz, 2016). Such differences in student preparation are not present in the heterogeneity analyses of college selectivity papers.

6. Results for later cohorts

This last section shows that the effects of the Colombian admission reform on the SES/exam score correlation and on assortativity persist beyond the initial cohorts who took the new exam. However, the negative effects on graduation rates and earnings in treated regions die out several cohorts after the exam reform—both at the market-level and for low SES students in particular.

The above results focus on the first two cohorts exposed to the new admission exam (2000–2001) for three reasons. First, the testing agency stopped collecting data on mother's education in subsequent cohorts. Second, my earnings records extend only through 2012,

[42] Appendix Figure A7 shows that the exam reform led to a substantial shift in the distribution of pre-reform academic ability at flagship colleges in treated regions.

so it is less likely that I observe individuals entering the labor force after college during my data period.[43] Third, students in later cohorts may have chosen different high schools once information on school-level scores on the new exam became available. This could alter the distribution of ability reflected in the socioeconomic index, X_i, which raises a potential threat to identification.[44]

Nonetheless, the evolution of reallocation effects across cohorts is informative about the mechanisms that underlie colleges' contributions to student human capital accumulation. In this final section I therefore examine the persistence of the above effects beyond the 2000–2001 cohorts. This involves addressing the three questions raised in Section 2.6 of the framework: 1) Did the exam overhaul's effects on the distribution of scores persist? 2) Were the effects on students' college choices persistent? and 3) Were the human capital outcomes persistent?

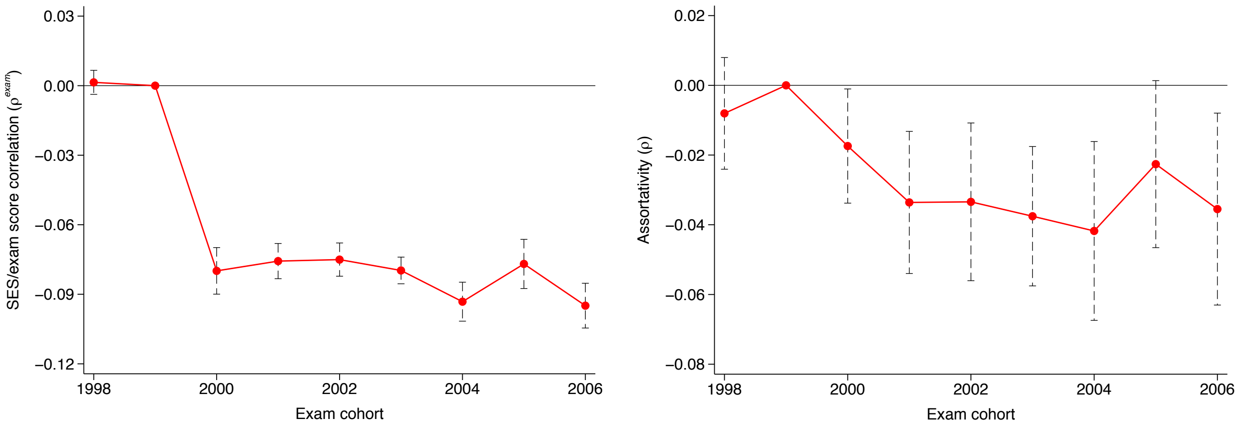
Figure 8 presents evidence related to the first two questions.[45] Panel A shows the correlation between the socioeconomic index, X_i, and admission scores for all exam takers in the 1998–2006 cohorts. This correlation is analogous to those in Figure 3, but I use only the average exam score across all exam subjects. Further, unlike in previous analyses, X_i is defined as the average 1998–1999 admission score for cells defined by only gender and high school; I do not also use mother's education to define X_i since I do not observe this variable beginning in 2003. The correlations reported in Figure 8 are relative to that for the 1999 cohort.[46]

[43] The time between high school graduation and college completion is much greater in Colombia than it is in the U.S. because students often do not enter college right away (Riehl, 2015). Further, most college programs take five years, and on-time completion rates are low.

[44] Students in the 2000–2001 cohorts were already in their final years of high school when the exam overhaul occurred, and thus it is unlikely that many students switched schools in response to the reform.

[45] Appendix Table A12 presents regression estimates corresponding to Figure 8.

[46] Specifically, the correlations in Panel A are estimated from a regression of exam scores on the socioeconomic index, X_i, interacted with cohort dummies, with the 1999 cohort as the omitted interaction. This regression also includes dummies for cohort-region cells and interactions of X_i with region dummies. In Figure 8, exam scores, X_i, and W_c are also normalized to mean zero and standard deviation one for each cohort. This ensures that the regression coefficients are equivalent to the correlation coefficients between these characteristics for each cohort.



Panel A. SES/exam score correlation

Panel B. Assortativity

FIGURE 8. Longer term exam score and college quality effects

Notes: Panel A shows results from a regression of average exam scores, τ_{it}, on dummies for cohort-region cells, interactions of the socioeconomic index, X_i, with region dummies, and interactions of X_i with cohort dummies. The panel plots coefficients from the interaction of X_i with cohort dummies with the 1999 cohort as the omitted interaction. Panel B plots event study coefficients β_t^ρ from equation (18), with the interaction between the treatment variable δ_{rt} and the dummy for 1999 graduates as the omitted group. Dashed lines are 90% confidence intervals using standard errors clustered at the region level. τ_{it}, W_c, and X_i are all normalized to mean zero and standard deviation one for each cohort.

There is little difference in the correlation of X_i and exam scores in the 1998 and 1999 cohorts, but this correlation falls sharply by roughly 0.1 points for the 2000–2001 cohorts. This replicates the results in Figure 3. However, Panel A also shows that the reform's effect on this correlation persists through the 2006 cohort with only small changes in its value. This suggests that the exam reform was successful in producing lasting changes in the distribution of exam scores, and that high SES students were not able to "undo" the reform's effects through exam prep or by choosing different high schools.

Panel B examines the persistence of the exam overhaul's effects on assortativity. This figure is similar to Panel A, but it reports the correlation of X_i with college quality, W_c, defined as each college's mean admission score for 1998–1999 enrollees. Further, Panel B is analogous to Figure 5 in that it reports event study coefficients from equation (18), which compares changes in this correlation between treated and control regions. Consistent with Figure 5, the change assortativity between 1998 and 1999 is similar in treated and control regions, and this correlation falls sharply for the 2000–2001 cohorts. However, Panel B also shows that, with the exception of the 2005 cohort, the decline in assortativity in treated

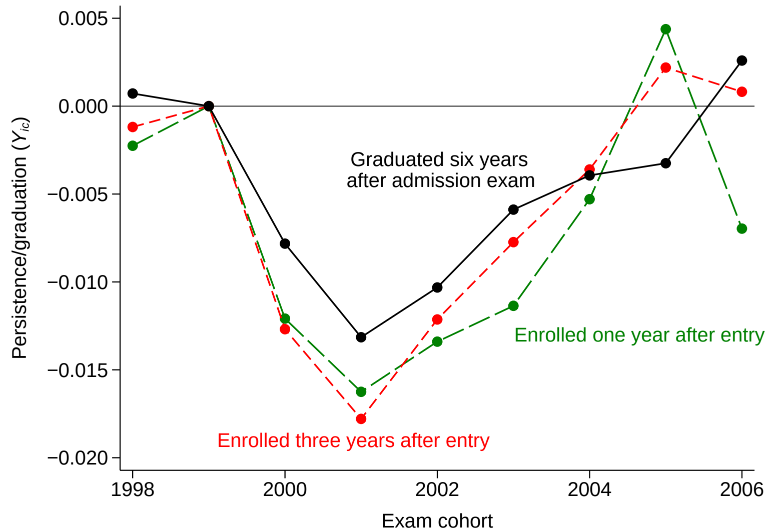


FIGURE 9. Longer-term persistence and graduation effects

Notes: This figure plots event study coefficients β_t from equation (19) for the three listed dependent variables, with the interaction between the treatment variable δ_{rt} and the dummy for 1999 graduates as the omitted group.

regions relative to control regions persists through 2006. Thus, the exam overhaul also appears to have had lasting effects on the distribution of student/college matches in treated regions. This is not surprising given the persistence of the exam score results in Panel A, since admission to flagship universities in treated regions was based purely on national exam scores.

The final question is whether the effects of the decline in assortativity on students' human capital accumulation are also persistent. Figure 9 analyzes students' shorter-term college success using outcomes that I can cleanly define given the time horizon of my data. Specifically, it examines whether students are still enrolled in college one year after entry, whether they are enrolled three years after entry, and whether they graduated college within six years of taking the admission exam. The graph displays event study coefficients analogous to Figure 6, with the 1999 cohort as the omitted group.

The pattern of coefficients is strikingly different from those in Figure 8. As in Figure 6, there are sharp declines in outcomes for the 2000–2001 cohorts in treated regions relative to control regions, but these decreases begin to fade out in the 2002 cohort. In the 2004–2006 cohorts the effects for all three outcomes are statistically indistinguishable from zero.

Thus while the exam overhaul appears to have generated a lasting effect on student/college assortativity, the negative effects of this reallocation on students' college outcomes appear to die out after several cohorts.

Table 7 reports regression estimates analogous to the results in Figure 9. The regressions for this table are similar to the benchmark differences in differences regression (19), but they include the 1998–2006 cohorts and calculate separate treatment effects for three post-reform cohort groups.[47] Columns (A)–(C) use the three outcomes from Figure 9 as dependent variables. The dependent variable for column (D) is log daily earnings. The regression in column (D) is more complicated because my earnings records cover only five years, which means I cannot measure earnings at the same time since the admission exam for all 1998–2006 cohorts. To address this, the regression in column (D) includes earnings measured at any "experience" level from six to 11 years after the admission exam, and it interacts all control variables with experience dummies. This ensures that the reported coefficients are identified only from variation in earnings with the same experience level.[48]

The top panel of Table 7 shows estimates for the full population of college enrollees. In the 2000–2001 cohorts, there are market-wide declines in persistence rates, graduation rates, and post-college earnings in treated regions relative to control regions. This is consistent with the findings in Table 6. However, these negative effects begin to die out in the 2002–2003 cohorts, and by the 2004–2006 cohorts the point estimates are near zero and not statistically significant.

[47] The full specification for this regression is:

$$Y_{irt} = \gamma_r + \gamma_t + \beta^{00-01}\delta_{r,00-01} + \beta^{02-03}\delta_{r,02-03} + \beta^{04-06}\delta_{r,04-06} + e_{irt},$$

where, for example, $\delta_{r,00-01}$ is an indicator equal to one for the 2000–2001 cohorts in treated regions.

[48] The full specification for column (D) is

$$Y_{irte} = \gamma_{re} + \gamma_{te} + \beta^{00-01}\delta_{r,00-01} + \beta^{02-03}\delta_{r,02-03} + \beta^{04-06}\delta_{r,04-06} + u_{irte},$$

where $e \in \{6\text{–}11\}$ denotes the number of years since the admission exam. Note that in this regression, treatment effects for the 2004–2006 cohorts are estimated using earnings measured six to eight years after the admission exam, when not all students had completed college.

Table 7. Persistence, graduation, and earnings in later cohorts

		(A)	(B)	(C)	(D)
		Dependent variable			
Treated regions × … × …		Enrolled one year after entry	Enrolled three years after entry	Graduated six years after exam	Log daily earnings
	2000−2001 cohorts	-0.013^{**}	-0.015^{**}	−0.011	-0.012^{**}
		(0.005)	(0.006)	(0.009)	(0.006)
All students	2002−2003 cohorts	−0.011	−0.009	−0.008	−0.011
		(0.008)	(0.008)	(0.011)	(0.010)
	2004−2006 cohorts	−0.001	0.000	−0.002	−0.004
		(0.013)	(0.011)	(0.010)	(0.011)
	2000−2001 cohorts	-0.015^{**}	-0.019^{***}	-0.018^{**}	−0.007
		(0.006)	(0.006)	(0.008)	(0.007)
X_i top quartile	2002−2003 cohorts	−0.011	−0.014	-0.021^{**}	−0.001
		(0.013)	(0.013)	(0.009)	(0.013)
	2004−2006 cohorts	−0.003	−0.008	-0.019^{***}	0.001
		(0.017)	(0.016)	(0.007)	(0.013)
	2000−2001 cohorts	-0.023^{**}	-0.024^{**}	−0.006	−0.005
		(0.009)	(0.009)	(0.009)	(0.010)
X_i bottom quartile	2002−2003 cohorts	-0.015^{*}	−0.005	0.008	−0.012
		(0.008)	(0.008)	(0.009)	(0.016)
	2004−2006 cohorts	−0.003	0.011	0.013	−0.002
		(0.012)	(0.011)	(0.009)	(0.015)
N		1,332,751	1,332,751	1,332,751	1,812,004
# regions		33	33	33	33
Dependent var. mean		0.649	0.554	0.111	10.294

Notes: The sample includes all college enrollees from the 1998–2006 exam cohorts. The top panel displays estimates of the β coefficients from the regression in footnote 47 for columns (A)–(C), and from the regression in footnote 48 for column (D). The bottom two panels display estimates from analogous regressions in which all control variables and coefficients are interacted with dummies for quartiles q of the socioeconomic index, X_i. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 1998–1999 cohorts.

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

The bottom two panels of Table 7 show analogous estimates for students in the top and bottom quartiles of the socioeconomic index, X_i.[49] For the 2000–2001 cohorts, the negative

[49] The estimates for both of these bottom panels come from a single regression that interacts the specification for the top panel with dummies for quartiles of X_i. To save space, I omit coefficients for the middle two X_i quartiles from Table 7; almost all of these coefficients are near zero and statistically insignificant.

effects on graduation and earnings outcomes mirror those in Table 6, although not all estimates are statistically significant.[50] With only one exception, these effects also fade toward zero in subsequent cohorts.

In sum, the results suggest that while the Colombian exam reform had a lasting impact on the distribution of test scores and the assortativity of student/college matches, its negative effects on graduation rates and earnings disappear after several cohorts. Six cohorts after the reform, there are no longer differences between treated and control regions in market-wide human capital measures. Further, there are few significant differences in these outcomes for either high SES or low SES students.

As discussed in Section 2.6, these findings suggest that time invariant mechanisms cannot fully explain how colleges affect student outcomes. For example, they are inconsistent with colleges providing an unchanging value added, or with graduation rates that depend purely on class rank. Instead, these results suggest that colleges can adapt to changes in their student bodies, perhaps by altering curricula or support services. Students may also respond to their new college types by increasing effort.

More broadly, these findings suggest that complementarity and mismatch are not fixed concepts. The negative effects for initially-exposed cohorts suggest that there can be adjustment costs to admission reform, but over time students or colleges may be able to take steps to address any resulting gaps in academic preparation. Further, my results suggest that findings in the college selectivity literature, which analyzes effects for marginal applicants, may not extend to settings with large-scale changes in the distribution of student/college matches. My results suggest that the nature of school quality can evolve with the composition of a college's student body.

[50] Unlike Table 6, not all of negative effects for the 2000–2001 cohorts in Table 7 are statistically significant. There are three reasons for this: 1) the dependent variable definitions are somewhat different; 2) I cannot use mother's education to define the socioeconomic index, X_i, for Table 7; and 3) I restrict the sample for Table 7 to students taking the admission exam in the year of their high school graduation.

7. Conclusion

A growing literature on college selectivity finds that, at least for certain individuals, attending a more selective college can lead to better career prospects (e.g., Dale and Krueger, 2002*a*; Hoekstra, 2009*a*). This work helps to explain why students and parents expend a great deal of energy on college admissions (Ramey and Ramey, 2010), as it addresses the salient question of whether college choice matters from an *individual's* perspective.

This paper has instead explored the consequences of college assignment from a *market-level* perspective. Specifically, it asked whether the matching of students to colleges can affect average outcomes across an entire college market. This follows in a long line of research on how the distribution of matches impacts aggregate outcomes (e.g., Gale and Shapley, 1962; Becker, 1973; Crawford and Knoer, 1981; Abdulkadiroğlu, Pathak and Roth, 2005).

A market-level focus raises issues distinct from those that are relevant to a college-bound individual. One must examine the full distribution of students and colleges, not just students on the margin of admission at selective colleges. Further, one must allow for general equilibrium effects of college choice; the selectivity literature reasonably assumes that college quality is unaffected by marginal changes in a school's admitted student body, but larger-scale changes in class composition may alter the mechanisms through which college quality affects outcomes. Finally, one may consider the possibility of an equity/efficiency trade off in college admissions (Durlauf, 2008); for example, does the allocation of selective college slots to disadvantaged students raise or lower the total productivity of a higher education system?

This paper has used data and a natural experiment from Colombia to provide evidence on the market-level consequences of the assignment of students to colleges. It analyzed a major reform of the national college admission exam in 2000, which led to higher scores for low SES students and thus relatively lower scores for high SES students. In certain regions, the reduction in the SES/exam score correlation led to market-wide decline in assortative matching on family and school resources, with some low SES students replacing high SES individuals at top colleges.

In the first two cohorts after the reform, market-wide graduation rates and post-college earnings fell in affected regions, with part of these declines coming from the lowest SES students who were shifted into higher quality colleges. These findings are consistent with a positive complementarity between socioeconomic status and college quality in the production of human capital outcomes, and they suggest that mismatch can arise in college assignment. This mirrors the finding in Arcidiacono, Aucejo and Hotz (2016) that science graduation rates would have been higher—for both minority students and for the University of California college system as a whole—in the absence of affirmative action policies. This paper extends these results to an entire college market and to labor market earnings after college completion. Further, it shows that this finding holds under weaker assumptions on the distribution unobservable match characteristics.

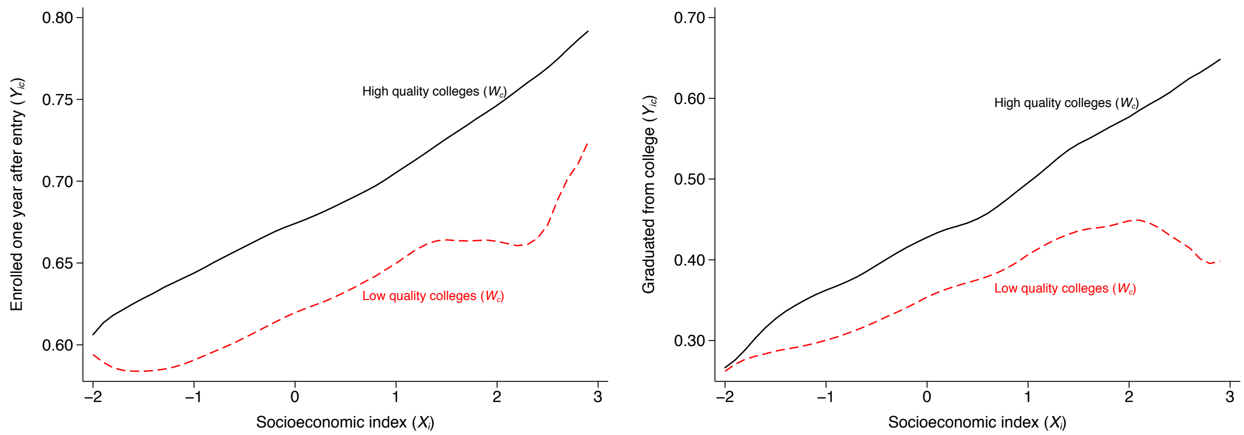
The results for initial cohorts suggest that college admission exams play an important role in matching students to schools at which they can succeed. Admission exams provide colleges with information on the types of students they are receiving, and thus changes in the structure of these exams can affect what colleges know about their student bodies. Reforms to admission systems, such as the 2016 SAT overhaul in the U.S. or the recent movement among top colleges to de-emphasize entrance exams, may therefore have consequences for both individual and system-wide outcomes.

However, this paper also showed that the negative effects of the Colombian reform on graduation rates and earnings died out after several exam cohorts. By the sixth post-reform cohort, there were no significant differences between affected and unaffected regions on market- or individual-level human capital outcomes, even while the reform's effects on assortativity persisted. Put another way, after several cohorts the return to college quality—for both high SES and low SES individuals—was zero.

The results for later cohorts suggest that the concepts of mismatch and complementarity are not set in stone. Colleges may have a comparative advantage in educating certain individuals, but these specializations may evolve to fit the backgrounds of their incoming students. For example, colleges may be able to change their curricula or provide more support services to help low SES individuals be successful. Such responses mean that the individual returns to college quality can differ from those that arise under large-scale admission system reforms.

The malleable nature of student/college complementarity weakens the argument that admission policies like affirmative action are undesirable because they create mismatch. My findings suggest that such reforms may come with adjustment costs but do not reduce the efficiency of a college system in the long run. Although I do not find evidence of equity gains in the form of higher graduation rates or earnings for low SES students, there may be other societal benefits to increased diversity (e.g., Rao, 2013; Fisman, Paravisini and Vig, 2016) that justify attempts to reduce the influence of socioeconomic status in college admissions.

A. Appendix



Panel A. Enrolled one year after entry

Panel B. Graduated from college

Figure A1. Observed complementarity in persistence and graduation rates

Notes: This figure is identical to Figure 1, except it uses different dependent variables. In Panel A, the dependent variable is a dummy equal to one if a student is still enrolled in college one year after entering. The dependent variable in Panel B is a dummy equal to one if a student graduated from any college.

A.1. **Complementarity on observable characteristics.** This section shows that the observed positive complementarity for log earnings documented in Section 2.5 also extends to other outcomes.

Figure A1 is analogous to Figure 1 in Section 2.5, but it uses different dependent variables. The dependent variable in Panel A is a dummy equal to one if a student is still enrolled in college one year after entering. The dependent variable in Panel B is a dummy equal to one if a student graduated from any college. The patterns in both panels are consistent with the positive complementarity observed for log earnings in Figure 1.

Table A1 presents the regression version of Figures 1 and A1. It displays the κ coefficients on the interaction term $X_i W_c$ from regressions using different outcome variables Y_{ic}. These regressions are similar to equation (10) in that they include controls for individual ability and college quality.[51] The dependent variables are the main outcomes that I analyze in Section 5: persistence in college—defined as being enrolled one year after entry (column

[51] Specifically, the regressions for Table A1 take the form $Y_{ic} = \gamma_{X_i} + \gamma_c + \kappa X_i W_c + e_{ic}$, where γ_{X_i} and γ_c are fixed effects for X_i cells and colleges. This specification provides an even stronger test of the null hypothesis $\kappa = 0$ than would a direct estimation of equation (10).

(A)), graduation from college (column (B)), and log daily earnings measures ten or eleven years after taking the Icfes exam (columns (C) and (D)). The κ coefficients are positive for all dependent variables and are statistically significant in all but one case.

Table A1. Estimates of pre-reform complementarity, κ_0

	(A)	(B)	(C)	(D)
	Dependent variable			
	Enrolled one year after entry	Graduated from college	Log daily earnings, 10 yrs later	Log daily earnings, 11 yrs later
Complementarity term ($X_i W_c$)	0.005*	0.006	0.024***	0.026***
	(0.003)	(0.004)	(0.004)	(0.004)
N	356,480	356,480	179,007	182,140
# colleges	296	296	267	268
Dependent var. mean	0.649	0.394	10.269	10.374

Notes: The table displays coefficients from a regression of the dependent variable listed in each column header on the interaction between the socioeconomic index, X_i, and observable college quality, W_c. See Appendix A.2 for details on the data and variable definitions. The regression also includes fixed effects for each college and each cell of the socioeconomic index. Parentheses contain standard errors clustered at the college level. Dependent variable means are calculated from the 1998–1999 cohorts.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2. **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 discusses how I select my *main analysis sample* for Sections 3–5, and my *longer-term analysis sample* for Section 6. Finally, it describes the 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. Between 2002–2006, the data include all students who took the exam in their last year of high school (11^{th} grade). In other words, the 2002–2006 data do not include individuals who repeated the exam or took it after they finished high school. For this reason, my main analysis sample includes all exam takers, but my longer-term analysis sample (which ranges from 1998–2006) includes only 11^{th} grade exam takers.[52]

[52] I identify 11^{th} grade exam takers in the 1998–2001 cohorts as those students who took the exam in their year of high school graduation.

Table A2. Construction of analysis samples

	(A) Main analysis sample (1998–2001 cohorts)	(B) Longer-term analysis sample (1998–2006 cohorts)
Total number of exam takers	2,100,424	4,269,762
Remove non 11^{th} graders	.	(559,826)
Missing exam scores	(10,195)	(2,699)
Missing high school information	(126,966)	(11,301)
Fewer than five pre-reform obs. in X_i	(84,464)	(448,473)
Full sample	1,878,799	3,247,463
College enrollees	714,071	1,332,751

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

My sample includes all exam takers with non-missing test scores whose high school identifier I observe in the data. In addition, I make a sample restriction that allows me to calculate a socioeconomic index, X_i, for each student. The index X_i is defined as the mean pre-reform exam score in cells defined by three characteristics:[53] 1) a dummy for a student's mother having a secondary education or above;[54] 2) dummies for a student's high school; and 3) a dummy for gender. I exclude students with values of X_i that are calculated based on fewer than five observations in the pre-reform cohorts (1998–1999).

Table A2 shows the effect of these restrictions on sample size. In both samples, the restrictions eliminate roughly ten percent of all exam takers in the ICFES records, although I also remove 559,826 non-11^{th} graders from my longer-term analysis sample from the 1998–2001 cohorts. The full sample includes 1,878,799 exam takers from the 1998–2001 cohorts for my main analysis, and 3,247,463 exam takers from the 1998–2006 cohorts for my longer-term analysis. However, most of my analyses are restricted to those who enrolled in college, as shown in the last row of Table A2. I test for and find no evidence of selection into college enrollment.

[53] This index is the average score across all exam subjects. I use a leave-one-out mean for each 1998–1999 exam taker so that X_i is not mechanically correlated with the outcomes examined below.
[54] Missing values of mother's education are counted as a value of zero for this dummy.

Table A3. 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.

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 entrance exam. The exit exam became a requirement for graduation from any higher education institution in 2009.

Column (A) in Table A1 the 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 normative program duration.[55] 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 of exam takers depicted in column (B). Column (C) shows that the Ministry of Education

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

records include all Universities but are missing a few colleges that provide more technical training.[56] Overall, 96 percent of exit exam takers attend colleges that appear in the Ministry of Education records.

I define my main measure of observable college quality, W_c, as the mean pre-reform exam score in each college c.[57] For the 20 colleges with fewer than ten pre-reform enrollees, I define W_c as the mean exam score 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 average daily earnings by dividing monthly earnings by the number of formal employment days in each month and averaging across the year. In addition, I test for and find no evidence of selection into formal sector employment.

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 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 broadly comparable to the higher education enrollment rate in Colombia during the same time period.[58] A better indicator of merge success is the percentage of college enrollees that appear

[56] 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*).

[57] This measure is the average score across all exam subjects.

[58] 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. Roughly 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.

in the admission exam records because all domestic college students must take the exam. Among enrollees who took the admission exam between 1998 and 2001, I match 88 percent.[59]

A.3. **The 2000 admission exam reform.** This section provides further details on the 2000 reform of the Colombian college admission exam and the exam subjects that were offered between 1998 and 2006.

The goal of the 2000 exam 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.

Figures A2–A5 present sample questions from the biology, language, math, and social sciences components of the pre-reform and post-reform exams. The sample questions reinforce 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 typically include a figure or passage that the student must interpret. Further, some pre-reform questions

[59] 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.

Panel A. Pre-reform sample question

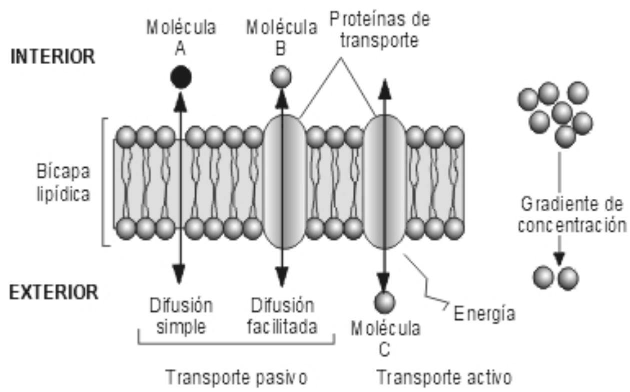
Which of the following two are inverse chemical processes?

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 A2. 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).

have a complicated answer structure, while the post-reform questions are all straightforward multiple choice.

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 A3. 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

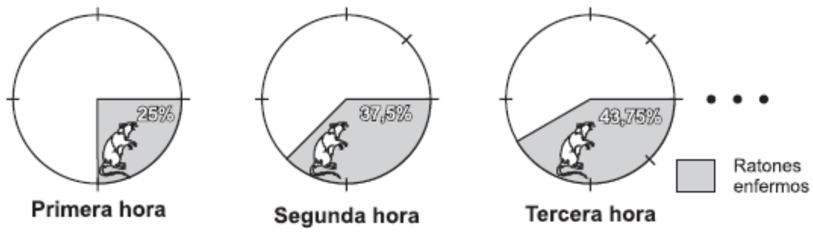
It is known that the result of multiplying a number by itself several times is 256. You can identify this number if it is known

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 A4. 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).

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 A4 shows the subject components that were included in the admission exam between 1998 and 2006. The 2000 reform combined two math exams—one designed to measure aptitude and another

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 A5. 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).

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 3 I focus on the six subject groups listed in the leftmost column of Table A4: 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 A4. Mean admission score by exam component and cohort

Subject groups	Exam components	1998	1999	2000	2001	2002	2003	2004	2005	2006
		Exam cohort								
Biology	Biology	47.6	48.1	45.1	44.6	45.0	45.3	46.0	47.3	47.0
Chemistry	Chemistry	46.1	51.0	45.0	45.2	44.3	43.5	42.3	43.6	45.2
Language	Language	48.6	50.9	46.5	46.4	48.3	48.9	52.4	46.4	48.4
Math	Math aptitude	49.1	50.5							
	Math knowledge	49.0	49.0							
	Math			43.0	41.1	42.7	41.8	41.0	44.5	45.7
Physics	Physics	47.3	47.1	45.3	46.7	45.3	46.2	42.9	46.7	45.9
S. sciences	Social sciences	47.9	48.6							44.9
	Geography			44.4	43.0	43.4	42.9	49.6	41.3	
	History			43.5	43.5	43.4	43.3	44.2	42.5	
Excluded components	Verbal aptitude	48.3								
	Philosophy			44.8	43.9	44.7	44.9	45.4	43.6	47.1
	Foreign language			41.0	42.3	42.1	41.7	39.6	43.4	43.1
	Elective	49.5	50.9	52.3	55.4	52.4	48.5	48.7	47.5	48.0
	Mean (all components)	48.2	49.5	45.1	45.2	45.2	44.7	45.2	44.7	46.2
	St. dev. (all components)	10.2	10.2	7.4	7.5	7.4	7.4	8.0	8.0	8.0

Notes: The sample includes only students who took the exam in the year of their high school graduation.

Table A4 also shows that the reform affected the mean and the variance of exam scores. The bottom rows show that the mean score across all subjects was approximately 50 in the pre-reform cohorts, and approximately 45 in the post-reform cohorts. Further, the standard deviation across all components fell from approximately ten to 7.5. My interest is in students' relative performance, and so for all my analysis I normalize exam scores to be mean zero and standard deviation one within each exam cohort.

A.4. **Characteristics measured by the post-reform exam.** Figure 3 in Section 3 shows that low SES students scored relatively higher on the post-reform exam. This result is consistent with the testing agency's goal to design an admission exam with less socioeconomic bias, though it raises an additional question: what characteristics did the new exam measure? Another objective of the reform was to design a better measure of the competencies that predict success in college, but the decrease in correlation of scores with SES could arise if the new exam simply contained more measurement error.

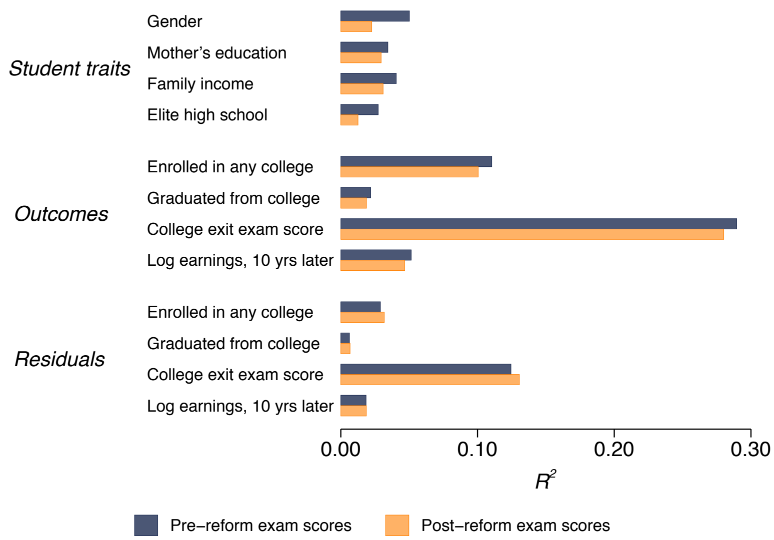


FIGURE A6. Explanatory power of the pre- and post-reform exam for students who took both

Notes: The sample includes students in my sample who took both the pre-reform exam and the post-reform exam once. Dark blue bars show the R^2 values from regressions of each listed dependent variable on linear terms for the pre-reform exam subject scores. Light orange bars show the R^2 values from regressions of each dependent variable on linear terms for the post-reform subject scores. The bottom block displays R^2 values from regressions in which the dependent variable is the residual from regressing each listed variable on indicators for mother's education, high school, and gender..

Although is hard to distinguish between these two possibilities without detailed measures of individual skills, Figure A6 presents suggestive evidence that the redesigned exam captured characteristics that matter for college success. For this figure I analyze the small fraction of students who took both the pre- and post-reform exams.[60] I then ask which exam scores explain more of the variance in these students' characteristics and college outcomes.

Specifically, the dark blue bars show the R^2 values from regressions of each listed dependent variable on linear terms for the pre-reform exam subject scores. The lighter orange bars depict regressions on linear terms for each post-reform subject score.[61] Note that the two regressions for each dependent variable include the same individuals, and thus variation in R^2 values reflects only changes in explained variance and not differences in the sample.[62]

[60] Approximately five percent of 1998–1999 exam takers in my sample also took the 2000–2001 admission exam.

[61] Formally, I compare the R^2 values from the regressions $y_i = \theta'[\text{Pre-reform scores}]_i + u_i$ and $y_i = \phi'[\text{Post-reform scores}]_i + v_i$.

[62] The population does vary *across* dependent variables due to missing values or conditional outcomes.

The top block of Figure A6 shows four characteristics related to students' gender, socioeconomic status, and high school. The pre-reform exam scores explain more of the variance in each of these student traits than the post-reform scores. This is consistent with the decline in the correlation of exam scores with socioeconomic characteristics evident in Figure 3.

The dependent variables in the middle block of Figure A6 measure four post-high-school outcomes: enrollment in college, graduation from college, scores on a national college *exit* exam, and labor market earnings ten years after taking the admission exam. Examining these outcomes gives a sense of whether the new exam is a better predictor of "success" in college, which was a stated intent of the exam overhaul. As with the observable characteristics, the pre-reform scores explain more of the variance in these outcomes than the new scores. This might seem to imply that the old exam does a better job at predicting college outcomes, but each of these outcomes also depends on students' backgrounds. Thus the decline in the correlation of exam scores with student traits is likely to affect the correlation of scores with student outcomes.

To address this issue, the last block of Figure A6 examines these four outcomes *net* of students' background characteristics. Specifically, it uses the residuals from regressions of each outcome variable on mother's education, high school, and gender dummies.[63] After netting out these socioeconomic characteristics, the post-reform exam scores explain more of the variance in college outcomes than the pre-reform scores.

The "horse race" depicted in Figure A6 is not a perfect test of the skills measured by the admission exams; in particular, the outcomes are potentially endogenous to the scores.[64] Further, the differences in R^2 values for the outcome residuals are small and not statistically different in most cases. Nonetheless, the combined evidence from Figures 3 and A6 is consistent with the major stated objectives of the exam overhaul: the redesigned exam led to

[63] These are the same characteristics used to define the socioeconomic index, X_i.

[64] For example, the new exam may be a better predictor of college outcomes because its scores played a larger role in determining where students went to college. However, if I examine students who repeated the old exam twice, or the new exam twice, the relationship between exam scores, student characteristics, and outcomes show different patterns from those in Figure A6. This suggests that the results in Figure A6 are driven by the admission exam reform rather than a repeat examination effect.

a dramatic reduction in the correlation of admission scores and SES without substantially reducing its predictive power for college success.

A.5. **Flagship universities and their admission methods.** Table A5 shows the flagship universities in Colombia, and it describes how I define treated and control regions. Column (A) lists the 33 administrative departments in Colombia that I call regions, and column (B) shows the average number of national exam takers per year in the 1998–1999 cohorts.

I define flagships as the largest public university in each region. "Largest" is defined by the total number of students for all cohorts in my enrollment records, and the restriction to "universities" excludes colleges that the Ministry of Education classifies as "university institutes" or "technical institutes." In addition, I make two modification to this definition of flagships. First, I consider Universidad Nacional de Colombia—which is the most prestigious public college in the country—to be the flagship school in Bogotá even though there are several other public universities with more enrollees.[65] Second, in the region of Norte Santander, I consider Universidad Francisco de Paula Santander to be the flagship since it is located in the capital city of Cúcuta, even though the public university in the city of Pamplona has more enrollees.[66] Column (C) lists the flagship universities in each region under this definition, and column (D) shows the cities in which they are located. In three regions, the largest public college is not a university (see column (E)), so I classify these regions as not treated. This classification captures the fact that non-university level colleges are typically open enrollment, and thus the exam reform is not likely to affect admission to these colleges.

Column (F) lists the admission method that each flagship used before the 2000 reform. I collected information on pre-reform admission methods 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. The majority of flagships used only

[65] Bogotá is the capital of Colombia and is its own administrative region.

[66] This does not affect the classification of Norte Sandander as a treated region as both colleges used only national exam scores for admissions.

Table A5. Flagship universities and definition of treated regions

(A) Region	(B) No. of exam takers	(C) College name	(D) City	(E) Type	(F) Admission method	(G) Closest other region	(H) Treated region?
Bogota	89,995	Universidad Nacional de Colombia	Bogota D.C.	University	Other		
Antioquia	61,338	Universidad de Antioquia	Medellin	University	Other		
Valle	54,760	Universidad del Valle	Cali	University	National exam		✓
Atlantico	28,814	Universidad del Atlantico	Barranquilla	University	Other		
Santander	26,167	Universidad Industrial de Santander	Bucaramanga	University	National exam		✓
Cundinamarca	24,486	Universidad de Cundinamarca	Fusagasuga	University	National exam		✓
Bolivar	18,340	Universidad de Cartagena	Cartagena	University	Other		
Boyaca	17,956	Universidad Pedagogica y Tecnologica de Colombia	Tunja	University	National exam		✓
Tolima	16,439	Universidad del Tolima	Ibague	University	National exam		✓
Cordoba	16,344	Universidad de Cordoba	Monteria	University	Other		
Narino	15,002	Universidad de Narino	Pasto	University	National exam		✓
Norte Santander	14,010	Universidad Francisco de Paula Santander	Cucuta	University	National exam		✓
Cauca	13,942	Universidad del Cauca	Popayan	University	National exam		✓
Caldas	12,637	Universidad de Caldas	Manizales	University	National exam		✓
Huila	11,986	Universidad Surcolombiana	Neiva	University	National exam		✓
Magdalena	11,462	Universidad del Magdalena	Santa Marta	University	Other		
Risaralda	11,064	Universidad Tecnologica de Pereira	Pereira	University	National exam		✓
Cesar	9,666	Universidad Popular del Cesar	Valledupar	University	National exam		✓
Meta	8,947	Universidad de Los Llanos	Villavicencio	University	National exam		✓
Sucre	7,472	Universidad de Sucre	Sincelejo	University	National exam		✓
Quindio	6,964	Universidad del Quindio	Armenia	University	National exam		✓
La Guajira	5,656	Universidad de la Guajira	Riohacha	University	Other		
Choco	3,532	Universidad Tecnologica del Chocodiego Luis Cordoba	Quibdo	University	Other		
Caqueta	3,020	Universidad de la Amazonia	Florencia	University	National exam		✓
Casanare	2,304	Fundacion Univ. Internacional del Tropico Americano	Yopal	Univ. Inst.			
Arauca	1,958					Norte Santander	✓
Putumayo	1,822	Inst. Tecnologico del Putumayo	Mocoa	Tech. Inst.			
San Andres	774	Inst. Nacional de Formacion Tech. Prof. de San Andres	San Andres	Tech. Inst.			
Amazonas	548					Caqueta	✓
Guaviare	304					Meta	✓
Vichada	164					Norte Santander	✓
Vaupes	114					Meta	✓
Guainia	91					Meta	✓

Notes: The number of exam takers is the average number per year calculated from the 1998–1999 cohorts.

national exam scores for admission. Eight flagships used other admission methods, which most commonly meant that they required applicants to take the university's own admission exam. In some cases these flagships also considered other information such as high school GPA or personal interviews.

Six regions of Colombia do not contain any colleges in my records. To classify treatment, I assign these regions to the closest region with a flagship using the distance between capital cities. Column (G) shows the closest assigned region for these six regions without any colleges.

Finally, column (H) of Table A5 shows my classification of regions as treated or control. Treated regions are those with flagship universities that use only the national exam for admissions, or those regions without colleges whose closest region has a national exam flagship. Control regions are those with flagship universities that use other admission methods, or those regions whose largest college is not a flagship. Summary statistics on the college markets and student populations in treated and control regions are in Table 1.

A.6. **Robustness tables.** The remainder of this appendix contains Tables A6–A12 and Figure A7, which present robustness and heterogeneity results as described in Sections 3–6.

Table A6. Exam reform effects on the SES/score correlation

Dependent variable: Normalized average exam score (τ_{it})

	(A)	(B)	(C)	(D)	(E)
		Alternative definitions of X_i			
Post-reform cohorts × ...	Main definition of X_i	Only mother's education	Only gender	Mean mother's edu. in HS	Only 1998 cohort
Socioeconomic index (X_i)	-0.080^{***}	-0.009^{**}	-0.038^{***}	-0.035^{***}	-0.079^{***}
	(0.007)	(0.004)	(0.004)	(0.006)	(0.006)
N	1,878,799	1,878,799	1,878,799	1,878,799	1,831,902
R^2	0.347	0.122	0.076	0.235	0.330
# regions	33	33	33	33	33
Mean SES/score correlation	0.617	0.258	0.147	0.450	0.594

Notes: This table reports results from a regression of average exam scores, τ_{it}, on an interaction of the socioeconomic index, X_i, with a dummy for the post-reform cohorts (2000–2001). The regression also includes dummies for cohort-region cells and interactions of the socioeconomic index with region dummies. Column (A) uses the benchmark definition of X_i, while columns (B)–(E) use the alternate definitions listed in each column header. Both $\tau_i t$ and X_i are normalized to mean zero and standard deviation one in all regressions. Parentheses contain standard errors clustered at the region level. Mean SES/score correlation is calculated from the 1998–1999 cohorts.

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

Table A7. Assortativity with major-level definition

	(A)	(B)	(C)	(D)
		Major definition		
Treated regions and cohorts (δ_{rt}) $\times$ ...	College level	Area (10 grp.)	Category (55 grp.)	Name (1000+ grp.)
Socioeconomic index $\times$ δ_{rt}	-0.025^{***}	0.004	0.005	-0.009
	(0.007)	(0.011)	(0.005)	(0.010)
N	714,071	714,071	714,071	714,071
# regions	33	33	33	33
Mean assortativity	0.429	0.205	0.271	0.408

Notes: This table reports results from regressions identical to those in Table 3, except they use different definitions of the dependent variable W_c. Column (A) in this table is identical to column (A) in Table 3. Columns (B)–(D) define W_c as the mean pre-reform exam score in different major categories as defined by the Ministry of Education. Parentheses contain standard errors clustered at the region level. Mean assortativity is calculated from the 1998–1999 cohorts.

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

Table A8. Formal labor market employment

	(A)	(B)
	Dependent variable	
Treated regions and cohorts (δ_{rt}) $\times$ ...	Employed in formal sector, 10 yrs later	Employed in formal sector, 11 yrs later
All students	−0.007	−0.011
	(0.008)	(0.008)
X_i top quartile	−0.017	−0.021
	(0.014)	(0.015)
X_i quartile 3	0.005	0.002
	(0.008)	(0.006)
X_i quartile 2	−0.001	−0.005
	(0.006)	(0.006)
X_i bottom quartile	−0.003	−0.014*
	(0.007)	(0.007)
N	714,071	714,071
# regions	33	33
Dependent var. mean	0.502	0.511

Notes: The top row reports estimates of β from equation (19). The bottom four rows report β_q coefficients from a version of (19) that is fully interacted with dummies for quartiles q of the socioeconomic index, X_i. The column header shows the dependent variable for each regression. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 1998–1999 cohorts.

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

Table A9. Clustered and wild t bootstrap p values

	(A)	(B)	(C)	(D)
	Dependent variable			
Treated regions and cohorts (δ_{rt}) $\times$ ...	Enrolled one year after entry	Graduated from college	Log daily earnings, 10 yrs later	Log daily earnings, 11 yrs later
All students	−0.014	−0.014	−0.014	−0.013
	(0.006)	(0.059)	(0.021)	(0.029)
	[0.036]	[0.082]	[0.030]	[0.054]
X_i top quartile	−0.017	−0.022	−0.013	−0.016
	(0.001)	(0.003)	(0.121)	(0.058)
	[0.028]	[0.016]	[0.102]	[0.096]
X_i quartile 3	−0.010	−0.004	−0.002	0.003
	(0.104)	(0.581)	(0.725)	(0.678)
	[0.140]	[0.632]	[0.750]	[0.718]
X_i quartile 2	−0.006	−0.001	0.001	0.001
	(0.382)	(0.898)	(0.861)	(0.911)
	[0.484]	[0.886]	[0.890]	[0.852]
X_i bottom quartile	−0.022	−0.022	−0.016	−0.013
	(0.007)	(0.018)	(0.063)	(0.027)
	[0.022]	[0.024]	[0.128]	[0.026]

Notes: This table presents regression coefficients identical to in Table 6. Parentheses contain p values from standard errors clustered at the region level, as reported in Table 6. Brackets contain p values from a wild t bootstrap imposing the null hypothesis, as recommended in Cameron, Gelbach and Miller (2008).

Table A10. Effects for top quartile of X_i by pre-reform college choices

	(A)	(B)	(C)	(D)	(E)	(F)
	Enrollment outcomes		Persistence, graduation, & earnings outcomes			
Treated regions and cohorts (δ_{rt}) × …	College quality (W_c)	X_i rank relative to class	Enrolled one year after entry	Graduated from college	Log daily earnings, 10 yrs later	Log daily earnings, 11 yrs later
Chose high quality	-0.059^{**}	0.012^{***}	-0.016^{**}	-0.035^{***}	-0.026^{**}	-0.023^{*}
	(0.024)	(0.003)	(0.006)	(0.006)	(0.010)	(0.011)
Chose low quality	-0.009	0.014^{**}	-0.018^{**}	-0.013	-0.005	-0.014
	(0.041)	(0.006)	(0.008)	(0.014)	(0.014)	(0.014)
N	292,209	292,209	292,209	292,209	153,648	157,132
R^2	0.080	0.074	0.004	0.008	0.041	0.038
# regions	32	32	32	32	32	32
Chose high mean	0.662	0.695	0.695	0.482	10.430	10.545
Chose low mean	0.191	0.790	0.679	0.448	10.358	10.470

Notes: The sample includes all college enrollees from the 1998–2001 exam cohorts in the top quartile of the socioeconomic index X_i. The regression displays coefficients from the standard differences in differences regression (19) using the dependent variable in each column header, but in which all control variables are interacted with a dummy for whether students are likely to choose top colleges or to choose lower-ranked colleges. "Chose high quality" means that the average value of college quality, W_c, within a student's X_i cell was greater than or equal to the predicted value of W_c from a local linear regression of W_c on X_i. "Chose low quality" means that the average value of W_c within a student's X_i cell was below this predicted value. Parentheses contain standard errors clustered at the region level. Dependent variable means for each student type are calculated from the 1998–1999 cohorts.

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

Table A11. Unexplained effects under double randomization

Treated regions and cohorts (δ_{rt}) × …	(A) Enrolled one year after entry	(B) Graduated from college	(C) Log daily earnings, 10 yrs later	(D) Log daily earnings, 11 yrs later
	Dependent variable: Total − distribution effect ($Y_{ic} - \hat{Y}_{ic}$)			
All students	-0.013^{***}	-0.012	-0.004	-0.003
	(0.005)	(0.007)	(0.005)	(0.006)
X_i top quartile	-0.016^{***}	-0.020^{***}	-0.003	-0.006
	(0.004)	(0.007)	(0.008)	(0.008)
X_i quartile 3	-0.010	-0.005	-0.004	0.001
	(0.007)	(0.009)	(0.006)	(0.008)
X_i quartile 2	-0.008	-0.004	0.002	0.003
	(0.007)	(0.011)	(0.006)	(0.006)
X_i bottom quartile	-0.023^{***}	-0.024^{**}	-0.016^{*}	-0.016^{**}
	(0.007)	(0.009)	(0.008)	(0.006)
N	714,071	714,071	354,034	361,044
# regions	33	33	33	33

Notes: The dependent variable for this regression is the difference between the observed outcome, Y_{ic}, listed in each column header and the predicted outcome, $\hat{Y}_{ic}$, as defined in Section 5.3. The top row reports estimates of β from equation (19). The bottom four rows report β_q coefficients from a version of (19) that is fully interacted with dummies for quartiles q of the socioeconomic index, X_i. Parentheses contain standard errors clustered at the region level.

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

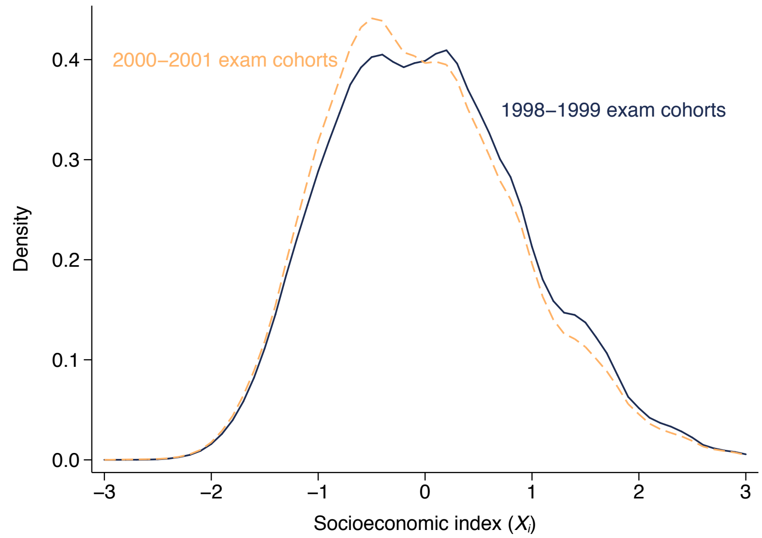


FIGURE A7. SES distribution in flagship universities with national exam admissions

Notes: This figure plots the distribution of the socioeconomic index, X_i, at flagship universities in treated regions. The dark solid line shows the distribution of X_i at these flagships in the two cohorts prior to the exam reform. The light-colored dashed line plots this distribution for the first two cohorts after the reform. 5.

Table A12. SES/exam score correlation and assortativity effects in later cohorts

	(A)	(B)
	Dependent variable	
	Average national exam score (τ_{it})	College quality (W_c)
2000−2001 cohorts	-0.079^{***}	-0.022^{***}
	(0.004)	(0.008)
2002−2003 cohorts	-0.078^{***}	-0.032^{**}
	(0.003)	(0.014)
2004−2006 cohorts	-0.089^{***}	-0.029^{*}
	(0.005)	(0.015)
N	3,247,463	1,332,751
R^2	0.294	0.181
# regions	33	33
Mean correlation	0.602	0.431

Notes: The sample for column (A) includes all 11^{th} grade exam takers from the 1998–2006 exam cohorts. This column reports coefficients from a regression of average exam scores, τ_{it}, on interactions of the socioeconomic index, X_i, with dummies for the three listed cohort groups: 2000–2001, 2002–2003, and 2004–2006. This regression also includes dummies for cohort-region cells and interactions of X_i with region dummies. The sample for column (B) includes all college enrollees from the 1998–2006 exam cohorts. This column reports coefficients from a regression analogous to specification (18), but in which the treatment variable is defined separately for the three cohort groups. Parentheses contain standard errors clustered at the region level. Mean correlations are calculated from the 1998–1999 cohorts.

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

Chapter 2. The Big Sort: College Reputation and Labor Market Outcomes

(with W. Bentley MacLeod, Juan E. Saavedra, and Miguel Urquiola)

1. Introduction

Each year, millions of high school graduates choose a college with the hope of a good career. The problem they face is daunting. Each potential college they consider may provide them with a different set of skills. Moreover, the labor market may use the identity of their college as a signal of ability. We call the process by which students choose a college, and find their first job, the "big sort." This paper provides evidence on the role of college identity in the big sort using unique data and a natural experiment from the country of Colombia.

We introduce a simple measure of a college's reputation: the mean admission score of its graduates. Our data allow us to observe which college individuals attended, as well as their subsequent performance in the labor market. We show that their earnings are positively correlated with the reputation of their colleges after controlling for individual characteristics, including their own admission scores. This correlation may arise because high-reputation colleges provide more skill, or because college identity signals graduates' ability. To differentiate between these mechanisms we exploit the staggered introduction of a national college *exit* exam that provided employers with a new signal of individual skill. A competitive labor market model predicts that the exit exam should reduce the correlation of earnings with college reputation if reputation serves to signal ability. The empirical evidence is consistent with this prediction, suggesting that college identity plays an informational role.

Finally, we measure the effect of college reputation upon subsequent earnings *growth*. We find that the correlation between reputation and log earnings is not constant, but rather increases with a worker's labor market experience. This descriptive result contrasts with a large literature on the Mincer wage equation, which finds that the correlation of log wages with workers' years of schooling does not vary with experience (Lemieux, 2006). Thus differences in educational attainment are an initial but stable source of inequality (Katz and Murphy, 1992; Autor, 2014). Our result shows that sorting across different types of colleges may be a further source of inequality—one that grows in importance over the course of workers' careers.

The paper proceeds as follows. In Section 2 we develop the model that guides our empirical analysis. The big sort is a complex process, and preferences over colleges may depend upon who else chooses to attend (Rothschild and White, 1995). Moreover, there are multiple sorting equilibria depending upon colleges' selection strategies (MacLeod and Urquiola, 2015). Rather than attempting to model this complexity, we build on the standard competitive model of wage formation (Burdett, 1978; Jovanovic, 1979).

We make two assumptions that yield clear predictions regarding the effect of school reputation on wages. First, colleges select students based on their scores on a standardized admission test, as is the case in Colombia. Hence the most desirable colleges tend to enroll the highest scoring students. This allows us to propose a simple definition of a college's reputation: the mean admission score of its graduates. Second, we assume employers observe graduates' college of graduation, but not their individual admission scores. Thus, in setting wages employers use college reputation to infer graduates' ability as measured by the admission test.

We explore the implications of these assumptions for the introduction of an individual-specific measure of skill that employers do observe—a college exit exam score. The key prediction is that in earnings regressions that include both college reputation and individual admission scores, the availability of the exit exam reduces the return to reputation and increases the return to individual scores.

In Section 3 we explore the empirical evidence on this prediction using administrative data that link, for all college graduates: scores on a standardized national admission exam, college of graduation, and labor market outcomes. During the period we study, Colombia introduced national college exit exams, and many students began listing their exit scores on their CVs. These exams were gradually rolled out across 55 fields of study such as accounting, dentistry, economics, and law. This allows us to implement an approach analogous to Card and Krueger (1992) who analyze how time-varying state policies (e.g., class size levels) affect a slope—the relation between years of schooling and wages. In our case the question is how time-varying college major characteristics (e.g., the existence of an exit exam in a related field) affect two slopes—the earnings return to reputation and the earnings return to admission scores. Consistent with the assumption that employers use college reputation to infer individual ability, we find that the new signal of skill reduced the return to reputation and increased the return to admission scores.

In addition, we find that the exit exams increased average earnings, a result that is consistent with improved employer-employee match quality.[67] The exit exams also prompted student behavioral responses in the form of delayed graduation and preference for colleges and programs with better exit exam performance. In short, these results provide evidence that college identity transmits information on ability, and that the reliance upon reputation fell in the presence of a better performance signal.

In Section 4 we ask whether college reputation relates to earnings exclusively through an informational channel. The competitive labor market model predicts that employers update their evaluation of a worker's skill based upon performance on the job (Harris and Holmstrom, 1982). Thus wages should become more correlated with ability as workers gain experience. Farber and Gibbons (1996) and Altonji and Pierret (2001) use this result to show that over workers' careers, observable characteristics like years of schooling become less correlated with wages in regressions that include unobserved measures of ability.[68] In our model, the fact that we define a college's reputation as the mean admission score of its graduates yields a clean prediction for regressions that also include individual scores. If reputation is solely a signal of ability as measured by admission scores, the correlation of earnings and reputation should decrease with experience conditional on individual scores.

The evidence is inconsistent with this prediction. We find that conditional on individual admission scores, the correlation between earnings and reputation *increases* with worker experience. This contrasts with Altonji and Pierret (2001), who find that the correlation

[67] Sahin et al. (2014) suggest that there is a role for policies that improve such matches. They point out that occupational mismatch in the U.S. has become more severe for college graduates since the great recession.

[68] Farber and Gibbons' and Altonji and Pierret's results suggest that schooling signals ability, while other factors correlated with schooling have a deterministic effect on wages. In related work, Lange (2007) finds that errors regarding worker skill decline markedly after a few years of employment, although Kahn and Lange (2014) find greater persistence.

of earnings and years of schooling decreases with experience conditional on measures of unobserved ability. This result suggests that college reputation may also affect earnings through channels other than signaling. We cannot disentangle which of several competing hypotheses might explain this finding. Students and parents likewise cannot measure the counterfactual effects of college reputation on earnings. However, they may observe the correlation of reputation and career prospects, which likely increases the demand for the most reputable colleges.

Our findings relate to four distinct literatures on: reputational markets, college choice, the impact of selective schools, and costly signaling.

1.0.1. *Reputational markets.* Nelson (1970*a*) introduced the idea that consumer goods are either inspection or experience goods. The quality of an inspection good can be determined before purchase; that of an experience good can only be determined after. Work in industrial organization (Melnik and Alm, 2002, Hubbard, 2002, Jin and Leslie, 2003, Cabral and Hortacsu, 2010, and Dranove, 2010) observes that with experience goods the reputation of the seller affects the price; for example, a bottle from a good winery commands a high price even if it ultimately proves to be corked. We show that a similar effect arises in education: employers are sensitive to college reputation, and this sensitivity is reduced when better information becomes available (as recommended by Bishop, 2004). Further, consistent with college being a complex, composite good (e.g., Black and Smith, 2006), we find that students in turn respond to employers' changing perception of college reputation.

1.0.2. *College choice.* Hoxby (1997, 2009) shows that stratification by ability has increased significantly among U.S colleges. Thorough sorting may account for the fact that Arcidiacono, Bayer and Hizmo (2010) find that college identity in the U.S. fully reveals Armed Forces Qualification Test (AFQT) scores. In contrast, we find that college identity only partially reveals admission scores in Colombia. This discrepancy may reflect that college stratification in Colombia, although increasing, is not as thorough as in the U.S.[69] This suggests that college preferences, and hence reputations, are endogenous and may change over time. In particular, the introduction of college exit exams affected the labor market return to college reputation and preferences of college applicants. The endogeneity of preferences is relevant to theoretical work on matching in college and other markets (e.g., Roth and Sotomayor, 1989; He, 2014).[70] These models assume that students have clear exogenous preferences over the colleges they wish to attend. Future research could explore if peer effects impact optimal market design.

1.0.3. *The effects of attending a selective college.* Our work complements studies that estimate the wage effects of attending a selective college. Using U.S. data, Dale and Krueger (2002, 2014) find a positive effect, but one that is concentrated among minorities (see also Hoekstra, 2009*b*). Using Chilean data, Hastings, Neilson and Zimmerman (2013*a*) find evidence of significant variation in effects across colleges and majors, and less heterogeneity across family background (see also Urzua et al., 2015). Our contribution is to explore the mechanisms underlying these effects by explicitly measuring reputation in an entire market. While our results suggest that information-related channels may account for some of the effects in this literature, they do not foreclose other mechanisms like peer effects (Epple et al., 2006) and network externalities (Zimmerman, 2013; Dustmann et al., 2016; Kaufmann, Messner and Solis, 2013). If such externalities are more important for high level managerial jobs, then this may explain the effect of college reputation upon wage growth.

1.0.4. *Costly signals.* The celebrated Spence (1973) model shows that if schooling is a signal whose cost is declining in ability, a college wage premium can exist even if college has no value added. Our focus is on *which* college students attend rather than *whether* they attend. As MacLeod and Urquiola (2015) show, students may care about college identity even if

[69] In addition, Hoxby and Avery (2013*a*) show that even controlling for ability, individuals from disadvantaged backgrounds are less likely to apply to reputable colleges. Their results are generally consistent with a role for "brand name" reputations.

[70] See Abdulkadiroglu and Sonmez (2013) for a recent review of the large literature on this issue.

college value added does not vary with college reputation. Informational concerns alone can lead to ability sorting and stratification.

Finally, such informational channels cannot explain our finding of a positive correlation between reputation and earnings growth. Greater traction might arise from models with peer effects (e.g., Epple and Romano, 1998a) especially if peer interactions provide networks that become more valuable with experience. Alternately, graduates from high reputation colleges might be more likely to obtain positions in firms with higher levels of on-the-job human capital investment. For example, larger firms sometimes set pay bands for positions they wish to fill with applicants of certain characteristics (this is known as the Hay compensation system; see Milkovich, Newman and Gerhart, 2011). Those characteristics might include college identity, leading to a correlation between reputation and investment in human capital that results in a superstar-type effect (Rosen, 1981).

2. College reputation, signaling, and wages

This section adds college reputation to the standard Bayesian model of wage formation (Jovanovic, 1979). It presents two propositions that we take to the data in Sections 3 and 4. A full derivation of the model and these propositions is in Appendix A.

2.1. **Ability, admission scores, and college reputation.** Let α_i denote the log ability of student i, where by ability we mean the type of aptitude measured by pre-college admission tests. We define two measures of ability from our data. First, we observe each student's score on a college admission exam, τ_i, and we assume it provides a noisy measure of ability:

$$\tau_i = \alpha_i + \epsilon_i^\tau.$$

The second measure is college reputation. Reputation may incorporate many aspects of college quality, such as peer composition and faculty research output. We define the reputation of a college s to be the mean admission score of its graduates, and denote it by

R_s:

$$R_s = E\{\tau_i | i \in s\} = \frac{1}{n_s} \sum_{i \in s} \tau_i,$$

where n_s is the number of graduates from college s. This measure has two analytical advantages. First, in settings where selective schools use test scores to determine admission, R_s will be mechanically related to other attributes that lead students to prefer certain colleges. Second, as we discuss below, this reputation measure delivers clear predictions in regressions that also include individual admission scores.

2.2. **Employers' information and wage setting process.** We let θ_i denote the log skill of student i and suppose it is given by:

$$\theta_i = \alpha_i + v_{s_i}.$$

Skill includes both pre-college ability, α_i, and v_{s_i}, which we will interpret as attributes related to an individual's membership at college s_i. These may include factors that contribute to skill formation at school, such as teaching or peer effects, as well as access to alumni networks. These may also include individual traits (not perfectly correlated with α_i) along which individuals sort into colleges, such as family income or motivation.

We suppose that the market sets log wages, w_{it}, equal to expected skill given available information, I_{it}, regarding worker i in period t:

$$w_{it} = E\{\theta_i | I_{it}\} + h_{it},$$

where h_{it} is time-varying human capital growth due to experience and on the job training. We consider Mincer wage equations that net out human capital growth to focus on the time-invariant component of skill that is generated by education and revealed over time to the employer (see Lemieux, 2006):

$$\hat{w}_{it} = w_{it} - h_{it} = E\{\theta_i | I_{it}\}.$$

We suppose that employers' information set, I_{it}, includes college reputation, R_{s_i}.[71] While employers likely care about individuals' pre-college ability as captured by R_{s_i}, they also care about other attributes related to graduates' post-college skill. We therefore define a college's *labor market reputation* as the expected skill of its graduates: $\mathscr{R}_s = E\{\theta_i | i \in s\}$. It follows that $\theta_{i \in s} \sim N(\mathscr{R}_{s_i}, \frac{1}{\rho^{\mathscr{R}}})$, where $\rho^{\mathscr{R}} = \frac{1}{\sigma^2_{\mathscr{R}}}$ denotes the precision of $\mathscr{R}_s$.[72]

Our data do not contain $\mathscr{R}_s$, and it may differ from R_s if colleges with higher reputation provide more value added or select students based upon dimensions of ability that we do not observe. For instance, if colleges prefer motivated students, and students prefer more value added, R_s and v_s will be positively correlated. To allow for this we suppose v_s satisfies $E\{v_s|R_s\} = v_0 + v_1 R_s$, where v_1 is the reputation premium, i.e., the return to reputation beyond that captured by admission scores. If this premium is positive ($v_1 > 0$) then a college with a better reputation provides higher value added, broadly understood.

To summarize, employers observe a signal of worker i's skill given by the labor market reputation of her college of origin:

$$\begin{aligned} \mathscr{R}_{s_i} &= E\{\alpha_i + v_{s_i} | R_{s_i}\} \\ &= E\{\alpha_i | R_{s_i}\} + v_0 + v_1 R_{s_i}. \end{aligned}$$

In words, labor market reputation captures employers' expectations of ability, α_i, and attributes related to college membership, v_s, under the assumption that they observe our measure of reputation, R_s.

At the time of hire, employers observe other signals of skill that we do not see (Farber and Gibbons, 1996). We denote these by:

$$y_i = \alpha_i + v_s + \epsilon_i,$$

[71] Employers likely observe college identity, but they may not perfectly observe our measure of reputation. Below we discuss how our definition helps to address the possibility that this assumption does not hold.

[72] We assume all variables are mean zero and normally distributed, and we characterize their variability using precisions. The precision, $\rho^{\mathscr{R}}$, could also be indexed by s and hence be school-specific. We did not find robust evidence that the variance has a clear effect on earnings, and so set this aside for further research.

with associated precision ρ^y. Importantly, y_i does not include τ_i because we assume that employers do not observe graduates' individual admission test scores. This is consistent with the standard assumption in the employer learning literature that AFQT scores are unobserved, and with anecdotal evidence that in our setting graduates' CVs rarely feature their college admission exam score (we present evidence supporting this assumption below).

Lastly, employers observe signals related to worker output after employment begins:

$$y_{it} = \alpha_i + v_s + \epsilon_{it},$$

where ϵ_{it} includes human capital growth and other fluctuations in worker output. These are observed *after* setting wages in each period t (where $t = 0$ is the year of graduation). Let $\bar{y}_{it} = \frac{1}{t+1} \sum_{k=0}^{t} y_{ik}$ denote mean worker output and let $\rho^{\bar{y}}$ be the time-invariant precision of y_{it}.[73]

The market's information set in period t is thus $I_{it} = \{\mathscr{R}_{s_i}, y_i, y_{i0}, ..., y_{i,t-1}\}$. Assuming all variables are normally distributed, log wages net of human capital growth are:

$$\hat{w}_{it} = \pi_t^{\mathscr{R}} \mathscr{R}_{s_i} + \pi_t^y y_i + \left(1 - \pi_t^{\mathscr{R}} - \pi_t^y\right) \bar{y}_{i,t-1}, \tag{1}$$

where the weights on the signals satisfy $\pi_t^{\mathscr{R}} = \frac{\rho^{\mathscr{R}}}{\rho^{\mathscr{R}} + \rho^y + t\rho^{\bar{y}}}$ and $\pi_t^y = \frac{\rho^y}{\rho^{\mathscr{R}} + \rho^y + t\rho^{\bar{y}}}$. Note that $\pi_t^{\mathscr{R}}, \pi_t^y \to 0$ as wages incorporate new information from worker output.

Equation (1) describes employers' wage setting process given available information, I_{it}. We do not observe I_{it}, and instead derive the implications of the wage equation for regressions on characteristics in our data. Below we estimate regressions that include controls for experience and graduation cohort to capture the time-varying effects (recall that $\hat{w}_{it} = w_{it} - h_{it}$). Here we focus upon the implications for the relationship between the signals of individual ability and wages net of human capital growth.

[73] The assumption that the precision of y_{it} is time stationary also follows Farber and Gibbons (1996). We note that this assumption implies that any human capital growth included in ϵ_{it} is not serially correlated.

We define the *return to reputation* at time t, r_t, and the *return to ability*, a_t, as the coefficients from the regression:

$$\hat{w}_{it} = r_t R_{s_i} + a_t \tau_i + e_{it}, \tag{2}$$

where e_{it} is the residual. The return to reputation, r_t, is the wage impact of a change in R_s for students with similar admission scores, τ_i. The return to ability, a_t, is the wage impact of a change in τ_i for students from colleges with similar reputations.

2.3. **Predictions for the introduction of a college exit exam.** While the returns to reputation and ability are not causal, changes in these parameters are informative as to the signaling role of reputation. In Section 3 we ask how these returns were affected by the introduction of a new measure of individual skill—a college exit exam. We suppose that the exit exam increases the amount of information contained in y_i; its precision is $\rho^{y,exit} > \rho^y$ when the exit exam is offered. This could arise because students list exit exam scores on their CVs, receive reference letters as a result of their performance, or modify job search behavior after learning their position in the national distribution of exam takers.

The increase in the precision of y_i reduces the weight on reputation in wage setting, $\pi_t^{\mathscr{R}}$. Let $\delta_i = 1$ if and only if a student is exposed to the possibility of writing the exit exam. We can rewrite regression (2) as follows:

$$\begin{aligned} \hat{w}_{it} &= (1-\delta_i)\left(r_t R_{s_i} + a_t \tau_i\right) + \delta_i \left(r_t^{exit} R_{s_i} + a_t^{exit} \tau_i\right) + e_{it}^{exit} \\ &= \left(r_t R_{s_i} + a_t \tau_i\right) + \delta_i \left(\beta_t^r R_{s_i} + \beta_t^a \tau_i\right) + e_{it}^{exit}, \end{aligned} \tag{3}$$

where $\beta_t^r = r_t^{exit} - r_t$ and $\beta_t^a = a_t^{exit} - a_t$. Appendix A.D shows that $\beta_t^r < 0$ and $\beta_t^a > 0$.[74] Thus we have:

[74] Since our regressions use log wages, the experience profiles reflect the reduction in uncertainty as information about the worker accumulates. Experience profiles can therefore differ for individuals with $d_i = 1$ and $\delta_i = 0$. To account for such effects, our regressions will include controls for experience that vary with individuals' potential access to the exit exams.

Proposition 1. If wages are set to expected skill given the available information (equation (1)), then the introduction of an exit exam reduces the return to college reputation ($\beta_t^r < 0$) and increases the return to ability ($\beta_t^a > 0$).

Proposition 1 yields a prediction regarding the role of college reputation in transmitting information on ability. If employers do not use reputation to set wages, a new signal of skill should have no effect on the relative weights of reputation and admission scores. If instead the exit exam causes employer to rely less on labor market reputation, $\mathscr{R}_s$, and more on other signals of worker skill, y_i, this reduces the effect of R_s (which is a better predictor of $\mathscr{R}_s$) and increases the effect of the admission score (which is a better predictor of y_i).

Though one could measure college reputation in many ways, our definition isolates a signaling mechanism because R_s contains no additional information on α_i given a student's individual score, τ_i. Proposition 1 thus captures how the introduction of new information shifts the weight in wage determination from the group to the individual level measure of ability. In contrast, other measures of reputation may be correlated with α_i even conditional on individual scores.

Our definition also helps distinguish a signaling channel from competing hypotheses such as accountability effects. For example, in our context there is anecdotal evidence of colleges adding test-preparation sessions after the exit exam introduction. The exit exams may also have prompted colleges to change their curricula, or students to work harder. Such changes would affect skill formation while at college, included in v_s; they would not affect pre-college ability, α_i, the focus of our analysis. Thus, while accountability-related responses can explain changes in the return to reputation, they cannot explain a shift in the weight from R_s to individual admission scores. We describe the empirical evidence on signaling and accountability effects in Section 3.

It is worth emphasizing that reputation as defined above is an *equilibrium* phenomenon (MacLeod and Urquiola, 2015). It depends upon the more desirable colleges selecting individuals based on their observed ability rather than on other factors. Such an equilibrium is self-enforcing in the sense that students have an interest in working hard to get into the best college. However, a consequence of this effect is that it makes it difficult for the market to observe the quality of education. This happens because the market only observes the overall skill of graduates, and thus cannot easily disentangle college value added from selection. Our results will provide some direct evidence of this effect.

2.4. **Predictions for wage growth.** In Section 4, we describe how the returns to reputation and ability change with experience, t, thereby comparing college reputation to other signals of ability studied in the literature. Previous research makes a distinction between *conditional* returns, given by equation (2), and *unconditional* returns, given by:

$$\hat{w}_{it} = r_t^u R_{s_i} + e_{it}^R \tag{4}$$

$$\hat{w}_{it} = a_t^u \tau_i + e_{it}^\tau. \tag{5}$$

The unconditional returns to reputation, r_t^u, and to ability, a_t^u, are the coefficients on reputation and the admission exam score in these separate regressions. In Appendix A.E we show that the evolution of the regression coefficients from (2), (4), and (5) satisfy Proposition 2:

Proposition 2. If wages are set equal to expected skill given the available information then:

(1) The unconditional return to reputation, r_t^u, does not change with experience.
(2) The unconditional return to ability, a_t^u, rises with experience.
(3) The conditional return to reputation, r_t, is smaller than the unconditional return, and with experience falls to v_1, the reputation premium.
(4) The conditional return to ability, a_t, is smaller than the unconditional return, and rises with experience.

Parts (1)-(2) of Proposition 2 mirror Farber and Gibbons' (1996) predictions that observable characteristics are fully incorporated in initial wages, while employers gradually learn about unobservable traits. Reputation, R_s, has a constant effect because it is observed at the time of hire, and signals from worker output, y_{it}, merely confirm employers' expectations. The effect of the admission score, τ_i, grows with experience because it is initially unobservable to employers and correlated with y_{it}.

Parts (3)-(4) predict a declining conditional return to reputation, and an increasing conditional return to ability. These match Altonji and Pierret's (2001) predictions for observable and unobservable characteristics, but our measure R_s makes for a clean test of the role of reputation in signaling. Since reputation is mean college admission score, τ_i is a sufficient statistic for ability, α_i, in regression (4). Thus, part (3) of Proposition 2 holds even if employers imperfectly observe R_s, or if α_i is correlated with human capital growth; all of these effects are captured in the admission score coefficients in (2).[75] The return to reputation should decline unless there is a time-varying effect of other college membership attributes, v_s, and these attributes are correlated with reputation ($v_1 > 0$).

Thus Proposition 2 allows us to explore whether the return to reputation arises solely because college identity signals ability as measured by admission scores. Rejection by the data would suggest that other college membership attributes lead reputation to be correlated with wage growth. We examine these hypotheses in Section 4.

3. The college exit exam

Proposition 1 provides predictions for the introduction of an exit exam under a competitive labor market model. This section explores the empirical evidence related to these predictions. We first discuss institutional background and our measure of reputation. We then turn to the exit exam, sample, empirical specifications, and results.

3.1. **Background and data sources.** Colombia's higher education system consists of public and private institutions that award various types of degrees. In this paper, we refer to "colleges" as institutions that award the equivalent of U.S. bachelor's degrees after four or

[75] The assumption that R_s contains no information for α_i conditional on τ_i may not hold if α_i affects individuals' choice of college beyond their admission scores. In this case, the optimal predictor of α_i could also include R_s since τ_i is a noisy measure of ability. We discuss this possibility in Section 4.3, where we explore sorting into colleges on other dimensions.

five years of study. Colombia also has institutions that specialize in two or three year degrees. We set these aside to focus on institutional identity within a single schooling level.[76]

To apply to college, students are required to take a standardized exam, the Icfes.[77] The Icfes is generally analogous to the SAT, but it is taken by the vast majority of high school seniors regardless of whether they intend to apply to college.[78] The Icfes plays a major role in college admissions: many schools extend admission offers based solely on students' Icfes performance; others consider additional factors, and a handful administer their own exams.

We use student names, birthdates, and national ID numbers to link individual-level administrative datasets from three sources:

(1) The Colombian Institute for Educational Evaluation provided scores for all high school seniors who took the Icfes between 1998 and 2012. It also provided college exit exam fields and scores for all exam takers in 2004–2011 (discussed below).

(2) The Ministry of Education provided enrollment and graduation records for students entering college between 1998 and 2012. These include enrollment date, graduation or dropout date, program of study, college, and aggregate percentile on the Icfes exam. These data cover roughly 90 percent of all college enrollees; the Ministry omits a number of smaller colleges due to poor and inconsistent reporting.

(3) The Ministry of Social Protection provided monthly earnings records for formal sector workers during 2008–2012. These come from data on contributions to pension and health insurance funds. We calculate average daily earnings by dividing base monthly earnings for pension contributions by the number of formal employment days in each

[76] The Ministry of Education classifies institutions into five types: universities, university institutes, technology schools, technology institutes, and technical/professional institutes; we define the first two as colleges. We also focus on the Ministry's "university-level" majors, which have normative durations of 4-5 years.

[77] Icfes stands for Institute for the Promotion of Higher Education, the former acronym for the agency that administers the exam. The agency is now the Colombian Institute for Educational Evaluation, and the exam is called Saber 11°. We use the name Icfes to match the designation during the period covered by our data.

[78] Angrist, Bettinger and Kremer (2006) and our personal communications with the Colombian Institute for Educational Evaluation suggest that more than 90 percent of high school seniors take the exam. The test-taking rate is high in part because the government uses Icfes exam results to evaluate high schools.

month and averaging across months.[79] This agency also provided four-digit economic activity codes for the first job in which a worker appears in their records.

3.2. **Ability and college reputation.** We define two measures of ability that correspond to those in the theory (Section 2). The first is student i's score on the Icfes admission exam, which we denote by τ_i. Throughout, we express Icfes scores as percentiles relative to all high school seniors who took the exam in the same year. The second is the reputation of a college s, denoted by R_s, defined as the mean Icfes score of its graduates.[80] To avoid capturing any effects from the exit exam rollout on reputation, we calculate R_s using graduates who took the Icfes exam in 2000–2003.

Icfes and reputation are divided by ten so that both measures range from 0–10 and one unit is ten percentile points. One unit of reputation is about one standard deviation in this measure, and it is roughly sufficient to move from either the 75^{th} to the 100^{th} percentile, or from the 50^{th} to the 75^{th}. Anecdotally, a student applying to a very top college might also apply to one with one point lower in reputation as a "safety school."

Figure 1 shows that there is substantial variation in ability both across and within colleges. The horizontal axis depicts the reputation of 136 colleges that have at least ten graduates per cohort. The height of the black dots indicates the median Icfes percentile among graduates from each school, while the vertical bars show 25^{th}–75^{th} percentile ranges. There is a mass of colleges near the middle of the reputation distribution and fewer near the extremes. In addition, graduates from the same college differ significantly in ability. For example, the interquartile range at the median institution is 32 percentile points, which extends beyond the mean Icfes values of more than 80 percent of all colleges.

[79] Our theoretical predictions are for log wages, but our records only allow us to calculate earnings per day, not per hour. Colombian labor market survey data shows that hours are relatively constant early in college graduates' careers, which suggests that our results are not due to the use of daily earnings.

[80] In Colombia, students apply not just to a college but to a college/major pair. We define reputation at the college level to focus on the signaling component of a student's choice of institution. Major choice may also convey information about a student's ability. Below we show that our main results are similar when we define reputation at the college/major level.

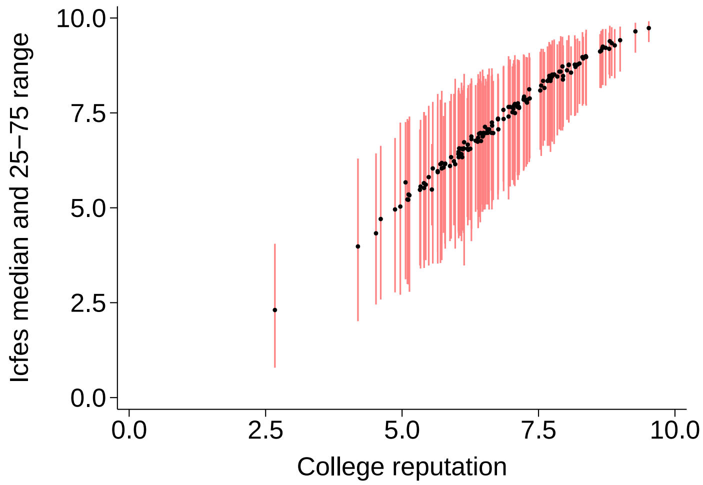


FIGURE 1. College reputation and individual ability

Notes: The sample for this figure includes all high seniors who took the Icfes in 2000–2003 and graduated from one of the 136 colleges with 40 or more graduates from the 2000–2003 Icfes cohorts (i.e., not less than ten per cohort). We define Icfes percentiles based on students' performance relative to all 11^{th} grade exam takers in their same year. Percentiles are calculated using the average of eight core component scores: biology, chemistry, geography, history, language, mathematics, philosophy, and physics. College reputation is the mean Icfes percentile among graduates from each of the 136 colleges. Black dots are the median Icfes percentiles among graduates from each school, and vertical lines are the 25^{th}–75^{th} Icfes percentile ranges.

3.3. **The exit exam.** In 2004 the agency that administers the Icfes test began another major initiative by introducing field-specific college *exit* exams. These exams are standardized and administered in every college that offers a related program. Exam fields range from relatively academic in orientation (e.g., economics and physics) to relatively professional (e.g., nursing and occupational therapy). The stated intent of this effort was to introduce elements of accountability into the college market. School-level aggregate scores were made available and used by news outlets as part of college rankings.

Rather than focus on its accountability dimension, we analyze the exit exam as potentially affecting students' capacity to signal their skill. This is consistent with anecdotal evidence that many students list exit exam scores on their CVs or on online profiles.[81] The exit exam

[81] It may be puzzling that, anecdotally, some students list their exit but not their Icfes exam scores on their CVs. One potential explanation is that the Icfes scores are more difficult to interpret. The Icfes exam yields scores on eight or more different subjects, and during the period we analyze the testing agency did not provide an aggregate score to students. By contrast, during the period of our analysis the exit exams yielded a single score in a subject related to a student's major.

may also affect faculty recommendations or students' search behavior after learning their position in the national distribution of exam takers.

3.4. **Identification.** To identify the effects of this new signal of skill, we exploit the gradual rollout of the exam fields in an "intent to treat" spirit. Exams were introduced in 55 fields between 2004 and 2007. The initial fields were those related to popular majors such as economics and industrial engineering; fields corresponding to less common degrees were introduced later (Appendix B.A lists all fields and their introduction year). During this time the exams were not required, although they were taken by the majority of students in related majors. In 2009, the exit exam became mandatory for graduation, and a "generic competency" exam was made available for majors without a corresponding field.

Although the exit exams were field-specific, during the period we study there was no formal system assigning college majors to exam fields. This match is necessary to determine which majors were treated. We therefore perform this assignment ourselves using the Ministry of Education's 54 major groups, which we label programs.[82] We assign each of the 54 programs to one of the 55 exam fields if the program name appears in the name of the field exam. We assign programs without matching names to the generic competency exam introduced in 2009. Appendices B.A and B.B describe this matching procedure and show that our main results are robust to several alternative matching methods.

Table 1 summarizes the resulting match. For each year it lists the number of matched programs and the program areas they originate in. Programs related to agronomy, business, education, and health received exam fields almost exclusively in 2004, while natural science programs did so in 2005. Programs related to fine arts had no corresponding field until the introduction of the generic exam in 2009. Some programs in engineering and social sciences received fields in 2004, while others had none up to 2009. Most of our identification comes

[82] These programs aggregate approximately 2,000 college major names that vary across and within schools. For instance, the Ministry might combine a major named Business Administration at one college with one labeled Business Management at another if it considers that these have similar content.

Table 1. Introduction of exit exam fields and matched college programs

Exit exam fields	Matched programs	Program area	College programs		
2004 fields	30	Agronomy	Agronomy	Animal husbandry	Veterinary medicine
		Business	Accounting	Administration	Economics
		Education	Education		
		Engineering	Agricultural eng.	Architecture	Chemical eng.
			Civil eng.	Electrical eng.	Electronic eng.
			Environmental eng.	Food eng.	Industrial eng.
			Livestock eng.	Mechanical eng.	Systems eng.
		Health	Bacteriology	Dentistry	Medicine
			Nursing	Nutrition	Optometry
			Physical therapy		
		Social sciences	Communication	Law	Psychology
			Sociology		
2005 fields	5	Natural sciences	Biology	Chemistry	Geology
			Math/statistics	Physics	
2006 fields	1	Health	Surgical tools		
2007 fields	1	Social sciences	Physical education		
2009 generic exam	17	Engineering	Administrative eng.	Biomedical eng.	Mining eng.
			Other eng.		
		Fine arts	Advertising	Design	Music
			Plastic/visual art	Representative art	Other fine arts
		Health	Public health		
		Social sciences	Anthropology	Geography/history	Language/literature
			Library science	Philosophy	Political science
Total	54				

Notes: This table displays the match of college programs to the exit exam field and generic exam years. Programs are the Ministry of Education's 54 core knowledge groups, which are further categorized into the listed eight program "areas." Appendix B.A lists the exam fields and details how we match them to programs.

from a comparison of 2004 programs and 2009 programs. Engineering and social science programs potentially provide a compelling comparison because they appear in both groups.

We define a binary treatment variable δ_{pc}, which equals one if students in program p and graduation cohort c had an available exit exam in the matched field. Because students typically take the exam one year before graduating, the first treated cohort is that which graduated one year after the introduction of the field assigned to its program.[83] For example, $\delta_{pc} = 1$ for psychology students who graduated in 2005 or later because the psychology field

[83] Across all cohorts in our sample, approximately 58 percent of test takers took the exam one year before graduation, 20 percent took it in the year of graduation, and 22 percent took it two or more years before.

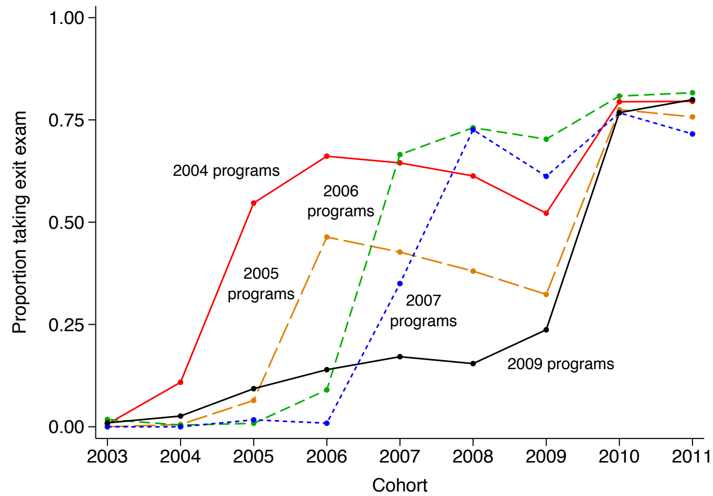


FIGURE 2. Proportion of students taking exit exam by program group

Notes: Lines represent program groups defined by the year in which the program's assigned exit exam field was introduced (see Table 1). The figure includes 2003–2011 graduates from all programs in our data, even those excluded from our main analysis sample for reasons described below.

exam was introduced in 2004. $\delta_{pc} = 0$ for all anthropology students who graduated before 2010 because the testing agency did not produce a related exam field.

Figure 2 shows that the introduction of exit exam fields led to sharp increases in the fraction of students taking the test. For example, the test taking rate in 2004 programs jumped from 10 to 55 percent with the 2005 cohort, the first we define as treated for this program group. Students in 2009 programs rarely took the exam until the cohort following the exit exam mandate in 2009.[84]

To summarize, we define a treatment indicator, δ_{pc}, at the program-cohort rather than at the individual level. Thus we analyze the introduction of the exams in an "intent to treat" spirit. This reflects that beyond the fact that students were not required to take exit exams during the period we study, they had no obligation to disclose their performance if they did (although not doing so might in itself convey information). Thus, while we can assert that the introduction of the exam into a student's field potentially affected the information

[84] The existence of exam takers in the 2003–2004 cohorts indicates that a small number of students took the exam in their final year or after graduating. The 75 percent test-taking rate in the 2010–2011 cohorts suggests that compliance with the exam mandate was not universal.

Table 2. Summary statistics for exit exam sample

	Year program received exit exam				
Variable	2004	2005	2006	2009	All
# graduates in 2003–2009	131,962	2,014	1,043	11,033	146,052
# earnings obs. in 2008–2012	528,435	7,418	4,516	41,433	581,802
# programs	27	1	1	10	39
# colleges	94	5	5	21	94
Reputation	7.45	8.50	5.88	8.26	7.52
	(1.21)	(0.66)	(0.42)	(0.96)	(1.21)
Icfes	7.66	9.04	6.36	8.60	7.74
	(2.29)	(1.09)	(2.27)	(1.71)	(2.26)
Log average daily earnings	10.87	10.71	10.66	10.84	10.87
	(0.70)	(0.66)	(0.51)	(0.76)	(0.70)
Return to reputation	0.138	0.041	-0.224	0.031	0.133
	(0.019)	(0.040)	(0.063)	(0.049)	(0.020)
Return to ability	0.028	0.009	0.015	0.049	0.029
	(0.003)	(0.010)	(0.014)	(0.013)	(0.003)

Notes: Log average daily earnings are for the year 2012. Parentheses contain standard deviations except for the returns to reputation and ability. These rows display coefficients on reputation and Icfes from a regression of log average daily earnings in 2008–2012 on these two variables, program-cohort dummies, and a quadratic in experience (defined as calendar year minus graduation cohort) interacted with program dummies. We run these regressions separately for each program group using only 2003–2004 graduates. The parentheses under these coefficients contain standard errors clustered at the college level.

available in that individual's labor market, we do not know precisely how it affected what firms observed about her.[85]

3.5. **Sample.** We analyze the effects of the exit exam using the 2003–2009 graduation cohorts. With these we can focus cleanly on the period in which signals of skill were introduced into a subset of fields.[86] Table 2 presents summary statistics separately for program groups defined by the year each program received its assigned exit exam field. Approximately 90 percent of students graduate from programs that received an exam field in 2004; most of the remaining graduates had no corresponding field until the 2009 generic exam.

[85] The potential endogeneity of exam taking also explains why we do not use the exit exam scores in our main analysis, either to define reputation or as a measure of graduates' skill. It is also possible that employers' perceptions of students from programs without exit exams were altered by information on programs in the same colleges that received exams. However, such spillover effects would bias our results toward finding no effects of the exit exams.

[86] This is no longer clearly the case after the 2009 cohort due to several structural changes in the exit exams.

We observe earnings for these graduates in 2008–2012. This means that we only observe earnings several years after graduation for cohorts prior to the exit exam introduction (2003–2004), while we observe earnings closer to graduation for cohorts after. The next section describes how we address this data constraint.

Our sample includes 39 programs offered at 94 colleges. These numbers are smaller than the total number of programs defined by the Ministry of Education (54) and the number of colleges in our records (136). We exclude programs and colleges that have too few observations to precisely estimate a return to reputation among graduates from the same program—a necessity for our empirical specification below. Appendices B.C and B.D provide details on the sample selection and show that our main results are robust to the key restrictions.

All colleges in the sample offer at least one of the 27 programs with a 2004 exam field, while only 25 schools offer one or more of the 12 programs with post-2004 programs. The distribution of Icfes scores is right-skewed with mean around the 77^{th} percentile—or 7.7 points. This reflects the fact that less than half of all high school graduates eventually enroll in college and, of those, about 50 percent graduate. Colleges that offer 2009 programs have reputations that are about eight percentile points higher on average than colleges that offer 2004 programs, but their graduates have slightly lower average daily earnings.

The last two rows in Table 2 report the returns to reputation and ability (Icfes) within each program group. These are analogous to the r and a coefficients from equation (2) in Section 2, except that these are averages across the multiple years of earnings we observe (2008–2012). In Table 2 we use only the two pre-exit exam cohorts (2003–2004) to estimate these returns; this provides a useful benchmark for the results below. 2004 programs have higher returns to reputation than the other program groups; a ten percentile increase in college reputation is associated with a 14 percent increase in earnings for 2004 programs, but only a three percent increase for 2009 programs.[87]

[87] The negative return to reputation for the 2006 program illustrates the empirical challenge of trying to estimate a return to reputation within each program. Not only can these returns be noisy when only a few schools offer a program, but the value of going to a higher-ranked school depends on the labor market that students from the program commonly enter (in this case, the program trains surgical instruments

These differences in program characteristics and returns raise questions as to whether delayed exit exam programs are a good counterfactual for early exit exam programs. We adopt several strategies to address these in our empirical analysis below.

3.6. **Empirical specifications and results.** This section estimates a benchmark specification that tests the effects of the exit exam on the returns to reputation and ability. We complement these results with four types of robustness checks. First, we add further controls for labor market experience and graduation cohort to address issues related to the structure of our data and to the years for which we observe earnings. Second, we restrict identification to programs with similar characteristics to address the non-random rollout of exam fields. Third, we explore the sensitivity of our results to competing hypotheses and other measures of college reputation. Fourth, we use balance and placebo regressions to test for differential sorting or concurrent macroeconomic trends.

3.6.1. *Benchmark specification.* We follow Card and Krueger (1992), who ask how state-level policies affect the rate of return to education. Note that the return to education is a slope—the impact of years of schooling on earnings. The issue we tackle is analogous—we ask how the exit exams affected the impacts of college reputation and Icfes on earnings. Our benchmark specification relates changes in the returns to reputation and ability to the staggered rollout of the exam fields. Consider the regression:

$$w_{ipct} = d_{pc} + f_p(t) + r_{pc}R_{s_i} + a_{pc}\tau_i + e_{ipct}, \tag{6}$$

where w_{ipct} is the log average daily earnings for student i in program p, graduation cohort c, and with potential labor market experience t, defined as calendar year minus graduation cohort. d_{pc} are dummies for program-cohort cells and $f_p(t)$ is a quadratic in experience interacted with program dummies. This "first-step" specification estimates returns to college reputation, r_{pc}, and to ability, a_{pc}, separately for each program-cohort cell.

technicians). For related issues see Hastings, Neilson and Zimmerman (2013*a*) and Urzua, Rodriguez and Reyes (2015).

A second-step regression relates these returns to our treatment variable δ_{pc}, which equals one for students with exit exam fields assigned to their program and cohort. For example, the second-step specification for the return to reputation is:

$$\hat{r}_{pc} = \mu_p + \mu_c + \beta^r \delta_{pc} + \upsilon_{pc}, \tag{7}$$

where μ_p and μ_c are program and cohort dummies and υ_{pc} is the residual. This is a standard differences in differences specification applied to slopes rather than to levels—it controls for average program and cohort differences in the returns to reputation (via the fixed effects μ_p and μ_c) and identifies the effect of the exit exam, β^r, through changes in returns across both programs and cohorts.

Card and Krueger (1992) use a two-step procedure. We opt for a single-step specification to identify changes in the *relative* weights of college reputation and Icfes on earnings. Plugging (7) and a similar equation for $\hat{a}_{pc}$ into (6) yields our benchmark specification:

$$w_{ipct} = d_{pc} + f_p(t) + (\mu_p + \mu_c + \beta^r \delta_{pc}) R_{s_i} + (\nu_p + \nu_c + \beta^a \delta_{pc}) \tau_i + e_{ipct}. \tag{8}$$

Specification (8) is analogous to equation (3) from Section 2, but it uses differences in differences variation in treatment. It controls for program-specific experience effects and level differences in daily earnings across program-cohort cells, and it allows each program and cohort to have different returns to reputation and Icfes through the μ and ν dummies. The coefficients of interest, β^r and β^a, are identified off variation in exposure to the exit exam across both programs and cohorts, defined by our treatment variable δ_{pc}.

Proposition 1 predicts $\beta^r < 0$ and $\beta^a > 0$. This comes from the assumption that employers use both labor market reputation, $\mathscr{R}_s$, and other signals of worker skill, y_i, in setting initial wages. We assume that the exit exam increases the precision of y_i, for example, through the appearance of scores on CVs. Our measure of reputation, R_s, is a better predictor of $\mathscr{R}_s$, while Icfes scores, τ_i, are a better predictor of y_i. Thus as the market relies less on $\mathscr{R}_s$ and more on y_i, the return to reputation falls ($\beta^r < 0$) and the return to ability rises ($\beta^a > 0$).[88]

Column (A) of Table 3 estimates benchmark specification (8). Like all other columns in Table 3 it reports only the β^r and β^a coefficients on the interactions of reputation and Icfes with our treatment variable δ_{pc}. The results suggest that relative to students in programs and cohorts without exams, students exposed to the exit exams see their daily earnings become more correlated with incoming collegiate ability and less correlated with college reputation. The reputation effect is slightly lower than one third of the mean return to reputation in Table 2; the Icfes coefficient is slightly higher than one half of the mean return to Icfes.[89]

Figure 3 illustrates the benchmark results in column (A) using only 2004 and 2009 programs. Panel A displays the linear relationship between reputation and residuals from a regression of log earnings on Icfes, experience, and program-cohort cells. The light-red lines depict programs with 2004 exit exam fields (Table 1) and the black lines contain programs that did not receive a field until 2009. In each case the solid lines describe students who graduated prior to the introduction of all exit exams, and the dashed lines students who graduated after the introduction of the initial exam fields. In 2004 programs, earnings are less correlated with reputation in cohorts following the exit exam introduction. In 2009 programs, the correlation between reputation and earnings is similar in all cohorts.

Panel B displays the analogous linear relationship between Icfes and log earnings residuals that control for reputation. The correlation between Icfes and earnings declines across cohorts in both program groups, but the decline is more pronounced in programs without an exam field. This is consistent with a stronger correlation between earnings and ability in early exit exam programs in the presence of an aggregate decline in the return to Icfes.

[88] Although this prediction results from higher precision in employers' initial information set, the changes in the relative returns to reputation and Icfes are also evident (but less pronounced) at periods $t > 0$ because wages continue to reflect initial information. Our data do not allow us to observe early career earnings for pre-exit exam cohorts (2003–2004), so our estimates reflect changes in returns at higher experience levels.

[89] Appendix B.E presents the program-cohort level returns to reputation and Icfes from the first-step equation (6). Averaging and differencing these returns yields estimates similar to column (A) of Table 3.

TABLE 3. Exit exam effects on returns to reputation and ability

Dependent variable: log average daily earnings

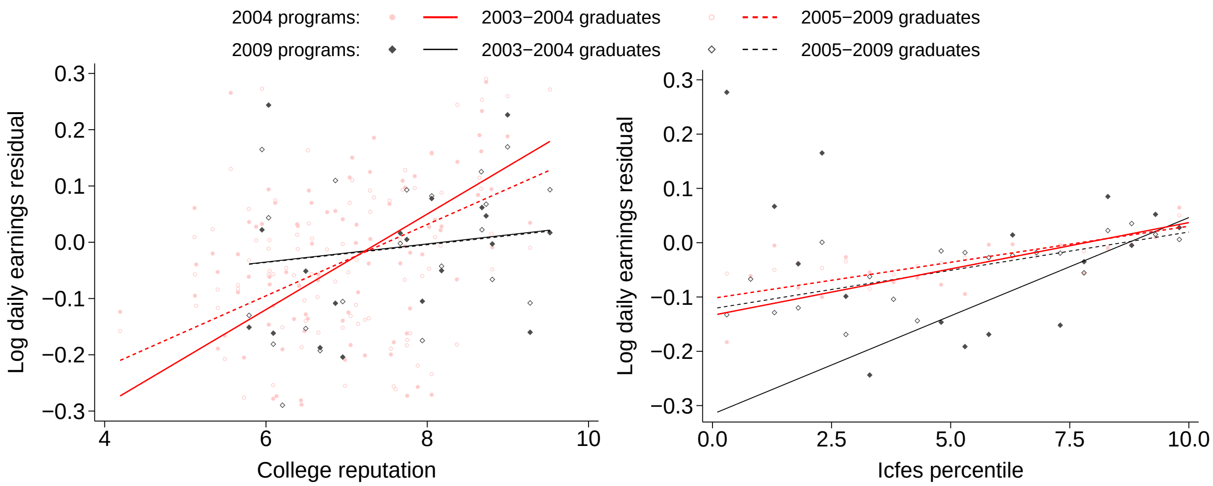
	(A)	(B)	(C)	(D)	(E)	(F)
		Experience & cohort controls		Restriction to similar programs		
	Benchmark specification	Within experience	Linear trends	S. sciences & engineering	Within $\hat{r}_p$ quartiles	Within $\hat{a}_p$ quartiles
Reputation × δ_{pc}	−0.041	−0.033	−0.034	−0.046	−0.018	−0.053
	(0.017)	(0.015)	(0.028)	(0.026)	(0.010)	(0.017)
Icfes × δ_{pc}	0.017	0.018	0.012	0.038	0.018	0.008
	(0.006)	(0.007)	(0.009)	(0.010)	(0.007)	(0.004)
N	581,802	267,924	267,924	273,590	581,802	581,802
R^2	0.258	0.224	0.224	0.266	0.258	0.258
# programs	39	39	39	22	39	39
Experience levels	0–9	4–7	4–7	0–9	0–9	0–9

Notes: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. Regressions in columns (A) and (C)-(F) include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. Column (B) includes dummies for program-cohort-experience cells and interactions of both reputation and Icfes with program-experience and cohort-experience dummies. The sample for each regression is restricted to the experience levels listed in the bottom row. Parentheses contain standard errors clustered at the program level.

Column (C) adds interactions of both linear experience and cohort terms with college reputation and Icfes for each program. Column (D) restricts the sample to social sciences and engineering program areas and adds interactions of dummies for social-science-area-cohort cells with both reputation and Icfes. Column (E) adds interactions of both reputation and Icfes with dummies for cells defined by cohort and each program's quartile of the returns to reputation estimated from 2003–2004 cohorts. Column (F) adds interactions of both reputation and Icfes with dummies for cells defined by cohort and each program's quartile of the returns to Icfes estimated from 2003–2004 cohorts.

There are two sources of caution in interpreting the results from (8)—one related to data constraints and one related to identification. The first arises because our data cover only seven cohorts with earnings observed over five years; hence we do not observe pre-treatment cohorts at very early experience levels. The second relates to possible violations of the usual assumption of parallel trends implicit in differences in differences estimation; evidence that such violations may be important comes from Table 2 and from the different pre-exit exam slopes in Figure 3. We now describe robustness checks that address these two issues.

3.6.2. *Experience and cohort controls.* Our sample includes 2003–2009 cohorts with earnings measured in 2008–2012. This means we cannot disentangle a first-period effect of the exit exam from an effect that varies with experience because we do not observe initial earnings for



Panel A. Return to reputation

Panel B. Return to ability

Figure 3. Exit exam effects—2004 and 2009 programs

Notes: In Panel A, the dependent variable is the residual from regressing log average daily earnings on Icfes, an experience quadratic interacted with program dummies, and program-cohort cell dummies separately for each program and cohort group. Lines depict the linear relationship between these earnings residuals and college reputation for each program and cohort group. Dots are the mean earnings residual at each college, calculated separately for each program and cohort group.

In Panel B, the dependent variable is the residual from regressing log average daily earnings on reputation, an experience quadratic interacted with program dummies, and program-cohort cell dummies separately for each program and cohort group. Lines depict the linear relationship between these earnings residuals and Icfes percentiles for each program and cohort group. Dots are the mean earnings residual in each of 20 equally-spaced Icfes percentile bins, calculated separately for each program and cohort group.

pre-exit exam cohorts. As a result, our benchmark results are based on returns to reputation and ability that average across experience levels.

Our data structure raises concerns if there is variation across programs in how college reputation or ability correlate with the returns to experience. For example, suppose that the return to reputation rises more quickly with experience in programs with early exit exam fields. This could mechanically generate a $\beta^r < 0$ estimate since the post-exam cohorts (2005–2009) have lower potential experience than the pre-exam cohorts (2003–2004).

To address this we add further controls for experience to the benchmark specification. To illustrate, suppose we estimated (8) using only earnings at five years of potential experience, thus ensuring that we are comparing exposed and unexposed cohorts at the same seniority. This regression could only include 2003–2007 cohorts because we do not observe earnings five years out for 2008–2009 graduates. We could repeat this estimation for any level of potential experience at which we observe cohorts prior to the introduction of all exit exams, which is between four (using 2004–2008 graduates) and seven (using 2003–2005 graduates) years of experience.[90] This procedure would yield four college reputation treatment effects and four Icfes treatment effects, one for each year of potential experience. We combine these into a single estimate by removing the experience quadratics from (8), restricting observations to those between four and seven years of experience, and fully interacting all fixed effects with experience dummies:

$$w_{ipct} = d_{pct} + (\mu_{pt} + \mu_{ct} + \beta^r \delta_{pc})R_{s_i} + (\nu_{pt} + \nu_{ct} + \beta^a \delta_{pc})\tau_i + e_{ipct}, \tag{9}$$

where d_{pct} are fixed effects for program-cohort-experience cells, and μ and ν are fixed effects for program-experience and cohort-experience cells. The coefficients β^r and β^a are thus averages of the experience-specific estimates, identified only off variation within experience levels. If unobserved program-level variation in the interaction of reputation and experience mechanically biases our estimate of β^r downward, including these experience controls should move the estimated coefficient toward zero.

The addition of experience controls decreases the magnitude of the reputation effect only slightly (Column (B), Table 3). Program differences in the returns to experience do not appear to drive the reduction in the return to reputation. This is also true for the return to Icfes; the estimates in columns (A) and (B) are nearly identical.

A related test is to allow the returns to reputation and ability to follow program-specific linear trends in both experience t and cohort c. For this we add linear trend interactions with reputation ($\mu_p t R_s$ and $\mu_p c R_s$) and with Icfes ($\nu_p t \tau_i$ and $\nu_p c \tau_i$) to the benchmark specification.[91] Including experience trends alone yields similar estimates to those from specification (9) since we limit the sample to earnings between four and seven years of experience. Adding

[90] In principle, we can identify treatment effects using post-2004 cohorts since two programs in our sample received the exit exam in 2005 and 2006. In practice, over 90 percent of our sample is comprised of students from 2004 programs, so regressions that exclude the 2003–2004 cohorts yield noisy estimates.

[91] The full specification with linear trends in experience and cohort is:

$$w_{ipct} = d_{pc} + f_p(t) + (\mu_p + \mu_p t + \mu_p c + \mu_c + \beta^r \delta_{pc})R_{s_i} + (\nu_p + \nu_p t + \nu_p c + \nu_c + \beta^a \delta_{pc})\tau_i + e_{ipct}.$$

cohort trends is the typical differences in differences test of adding linear terms in the "time" dimension. Cohort trends absorb linear program-specific paths in the returns to reputation and ability that predate the exit exam and should have a measurable impact on our point estimates if these paths are important.[92]

The results appear in column (C) of Table 3. The coefficient on the reputation effect is nearly identical to column (B), while the Icfes effect falls only slightly. The consistency of these magnitudes argues against the hypothesis of divergent trends across programs, although the estimates in column (C) are substantially less precise. This loss in precision suggests the effects of exit exam were not immediate but rather materialized over several years—an intuitive result if the market processed the tests gradually.

3.6.3. *Restriction to similar programs.* Our key identifying assumption is that in the absence of the exit exams, there would have been parallel trends in the returns to reputation and ability among programs exposed and not exposed to the exams. One fact that might cast doubt on this is that programs that received exams early have higher returns to reputation (Table 2). To address this we focus on comparable programs. We do so in three ways: i) restricting attention to social sciences and engineering, areas that have multiple programs in different exam year groups (see Table 1);[93] ii) stratifying programs by quartiles of the pre-exit exam returns to reputation, and iii) stratifying programs by quartiles of the pre-exam returns to Icfes. In each case we define program groups G and supplement equation (8) with dummies for group-cohort cells interacted with reputation and Icfes (e.g., $\mu_{Gc}R_s$ and $\nu_{Gc}\tau_i$).[94] Thus, β^r and β^a are only identified by variation in exposure to the exit exam within groups of programs that have common characteristics.

[92] Our ability to control for pre-existing cohort trends is limited, however, because we only observe two cohorts prior to the exit exam introduction (2003–2004).
[93] The health program area also includes a single program with a delayed exit exam field (surgical tools). Estimates analogous to column (D) that include health programs yield similar coefficients, but they are not significant because identification in the health program area comes from this single program.
[94] The full specification with program group controls is:

$$w_{ipct} = d_{pc} + f_p(t) + (\mu_p + \mu_c + \mu_{Gc} + \beta^r \delta_{pc})R_{s_i} + (\nu_p + \nu_c + \nu_{Gc} + \beta^a \delta_{pc})\tau_i + e_{ipct}.$$

Column (D) in Table 3 uses only programs in social sciences and engineering. The reputation effect is similar in magnitude to those in previous columns, while the Icfes effect is more than double. Both are statistically significant at the ten percent level despite the fact that the program restriction substantially reduces precision.

In column (E) we define program groups by pre-exit exam returns to reputation. We first estimate a return to reputation for each of the 39 programs in our sample using 2003–2004 graduates (i.e., $\hat{r}_{p,2003-2004}$).[95] We then define program groups G by quartiles of these returns, with 9–10 programs per group. This directly addresses the concern that 2004 programs have higher returns to reputation—in this case we compare delayed exam programs with low reputation returns only to the subset of 2004 programs with similarly low returns. The reputation effect in column (E) is smaller than in earlier specifications, consistent with some inflation in our estimates due to pre-treatment differences; but it is still significant because the standard error decreases. This suggests that the effects in this specification are identified off more similar programs because there is less noise in estimating treatment effects.

Column (F) is similar to column (E), but we define program groups as quartiles of pre-exit exam returns to Icfes (i.e., $\hat{a}_{p,2003-2004}$). This specification tests the influence of pre-treatment program differences in returns to ability. The resulting Icfes effect is also smaller than in the benchmark specification but more precisely estimated.

3.6.4. *Competing hypotheses.* The results in Table 3 are consistent with the exit exams transmitting information on ability, but the exams may have had other non-informational consequences. For example, school-mean exit exam scores were publicized, which may have altered employers' perceptions of colleges' labor market reputations. The exit exams may also have prompted colleges to change curricula or add test-preparation sessions, or individuals to work harder in preparation for the exams. Such accountability-related reforms may also have affected the observed returns to reputation and ability.

[95] For this we estimate equation (6) using only 2003–2004 graduates and replace the r_{pc} and a_{pc} coefficients with r_p and a_p. Appendix B.E presents these program-specific returns to reputation (and returns to ability).

In Appendix A.F we develop theoretical predictions that help distinguish between the informational and accountability-related impacts of the exit exams. The key insight is that accountability responses affect individuals' skill accumulation while in college—which is captured by the v_s term in our model—and not their pre-college ability, α_i. This generates different predictions as to how the exam introduction affects the returns to college reputation and to ability.

For example, in regressions that include both R_s and τ_i, an informational mechanism predicts a decrease in the return to reputation and an increase in the return to Icfes scores. However, accountability mechanisms predict no effect on the conditional return to τ_i because R_s is a better measure of college membership attributes, v_s. Thus while accountability responses could potentially explain the reputation effects in Table 3, they cannot explain the Icfes effects.

Further, signaling and accountability mechanisms generate different predictions regarding the *unconditional* returns to R_s and τ_i—that is, the coefficients in regressions that include only one of these characteristics. Under a signaling hypothesis, the exit exams should have no effect on the unconditional return to reputation if R_s is observed by employers, since information from individuals' exit exam scores merely confirms the labor market's expectations on average. Further, the exit exams should lead to an increase in the unconditional return to τ_i, but this increase should be smaller than when one also includes R_s in the regressions. By contrast, the main prediction from an accountability hypothesis is that the unconditional returns to R_s and to τ_i should have the same sign; this arises because any effects of the exams on v_s that are correlated with R_s will also, by definition, be correlated with τ_i.

Columns (A) and (B) in Table 4 present these unconditional returns to reputation and to ability by replicating our benchmark specification with only reputation terms (column (A)) or only Icfes terms (column (B)) included. As predicted by the signaling hypothesis, the unconditional return to reputation is not statistically different from zero, and the unconditional return to ability is positive but smaller than the conditional return (column (A), Table 3). Further, the unconditional returns to reputation and ability are oppositely signed.

Table 4. Exit exam effects under other reputation measures and hypotheses

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)
	Unconditional returns		Other reputation measures		
	R_s only	Icfes only	Mean Icfes at college-program	1 − admit rate at college	Mean log earnings at college
Reputation × δ_{pc}	−0.029		−0.038	−0.122	−0.044
	(0.018)		(0.022)	(0.066)	(0.080)
Icfes × δ_{pc}		0.005	0.019	0.012	0.012
		(0.007)	(0.007)	(0.006)	(0.006)
N	581,802	581,802	581,802	581,802	581,802
R^2	0.253	0.231	0.258	0.236	0.274
# programs	39	39	39	39	39
Mean return to reputation	0.161		0.132	0.098	0.700
Mean return to ability		0.069	0.027	0.064	0.035

Notes: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. Parentheses contain standard errors clustered at the program level.

Regressions in columns (A)-(B) are identical to column (A) in Table 3, except column (A) excludes Icfes and all its interaction terms, and column (B) excludes reputation and all its interaction terms.

Regressions in columns (C)-(E) are identical to column (A) in Table 3, except they use different reputation measures. Column (C) defines reputation as in our benchmark procedure (i.e., mean Icfes), but at the college-program level rather than the college level. Column (D) defines reputation as one minus the college admission rate (i.e., 1− admitted/applied) using aggregate admission data from the Ministry of Education. We include only university-level programs with a positive number of applicants and admitted students in a given cohort, and we average across all cohorts for which we have data (2007–2013). Column (E) defines reputation as the mean log daily earnings at each college using 2003–2004 graduates in our sample. We include only earnings at five years of potential experience, the earliest we can observe for both cohorts.

The mean returns to reputation and ability are calculated from specifications similar to those in each column, but they use only 2003–2004 graduates and include only a single reputation and Icfes term.

Thus, although the standard errors on these unconditional returns are too large to draw definitive conclusions, the results are more consistent with a signaling mechanism than with an accountability-related mechanism.

Taken together, the effects of the exit exams on the conditional and unconditional returns show that the strongest empirical result is the shift in weight from a group-level measure of ability—reputation—to an individual measure—Icfes scores. This is the effect captured by our benchmark specification, and it is harder to explain through channels other than signaling.

3.6.5. *Other reputation measures.* Our measure of reputation, R_s, captures the expected "admission exam" ability of graduates from a given college. The exit exams may also have provided information to employers on other dimensions of graduates' skill. Table 4 explores some of these. Columns (C)-(E) present results that use different measures of college reputation but are otherwise identical to our benchmark specification (Table 3, column (A)).

Column (C) defines reputation as mean Icfes at the college-program level rather than the college level, which allows schools to have strengths that vary by major. This is relevant because Colombian students apply to college/major pairs. We use a school-level definition for our main analysis to focus on the information conveyed by a student's choice of institution, but major choice may provide additional information on an individual's ability. The magnitudes of the results in column (C) are nearly identical to those for our benchmark results, though the standard errors are larger. This likely reflects the fact that the college-program reputations are calculated from smaller samples.[96]

Column (D) defines a college's reputation as one minus its admission rate (this measure is thus positively correlated with R_s). The results mirror those in our benchmark specification. The similarity of these results reflects the fact that R_s is mechanically correlated with other desirable school attributes when colleges use admission scores to select students.

Column (E) defines reputation as the average log earnings of a college's graduates.[97] This yields our best measure of labor market reputation, $\mathscr{R}_s$, which includes both pre-college ability, α_i, and attributes related to college membership, v_s. The exit exams led to an increasing return to Icfes and a lower return to reputation, though the reputation effect is statistically insignificant. Reputation measures like average earnings do not provide a clean test of signaling, however, because they may be correlated with ability even conditional on individual Icfes scores.

[96] Appendix B.L replicates all the robustness tests in Table 3 using college-program level reputation. The point estimates for both the reputation effects and the Icfes effects are similar to our main results across all specifications. The standard errors are typically larger, however, and thus the reputation effects corresponding to columns (B) and (D) of Table 3 are not statistically significant.

[97] We calculate this using only pre-exit exam cohorts (2003–2004) and earnings measured five years after graduation, the earliest we can observe for these cohorts. Results are similar when we use earnings measured in the year of graduation for cohorts exposed to the exit exams.

A related alternative hypothesis is that the exit exams enhanced the transmission of information on characteristics other than college reputation. This could explain the pattern of results in Table 3 if these characteristics are correlated with college reputation. To explore this hypothesis, in Appendix B.F we replicate our benchmark regressions including other individual characteristics—gender, mother's education, and family income—instead of reputation. These alternative specifications show that the returns to other characteristics also fell with the exit exams, but none of the results are statistically significant. Further, in a specification that includes all characteristics jointly, only the reputation effects are significant. Although we cannot rule out signaling effects on characteristics not included in our data, this provides evidence that the strongest effects of the exit exams were on the return to college reputation.

3.6.6. *Placebo and balance tests.* A further placebo test replicates our main analysis using college *drop-outs* rather than graduates. Drop-outs are a compelling placebo group because they enrolled in the same colleges and programs as graduates but exhibited little change in exam taking. Columns (A) and (B) in Table 5 document this by regressing an indicator for taking the exit exam on program dummies, cohort dummies, and our treatment variable, δ_{pc}. For graduates, exposure to the exit exam is associated with a 50 percentage point increase in the likelihood of taking the exam; for drop-outs it is unrelated.

Column (C) replicates our benchmark result for graduates from specification (8) (Table 3, column (A)). Column (D) estimates the same specification using drop-outs. There is little evidence that changes in drop-outs' returns to reputation and ability are correlated with the introduction of the exit exams. If anything, the return to reputation for drop-outs increases with the exam rollout, although the coefficient is noisily estimated. The point estimate on the Icfes effect is close to zero. To the extent that drop-outs and graduates are subject to similar enrollment or macroeconomic trends, this finding supports the notion that our main results are attributable to the exit exams.

Table 5. Placebo test using college drop-outs

	(A)	(B)	(C)	(D)
	Dependent variable: Took the exit exam		Dependent variable: Log average daily earnings	
	Graduates	Drop-outs	Graduates	Drop-outs
Exposed to exit exam (δ_{pc})	0.500	0.025		
	(0.054)	(0.020)		
Reputation × δ_{pc}			−0.041	0.011
			(0.017)	(0.032)
Icfes × δ_{pc}			0.017	−0.002
			(0.006)	(0.011)
N	146,052	77,586	581,802	259,258
R^2	0.335	0.026	0.258	0.118
# programs	39	39	39	39
Mean exam taking rate	0.070	0.162		
Mean return to reputation			0.133	0.040
Mean return to ability			0.029	0.032

Notes: The sample for columns (A) and (C) includes college graduates and their earning observations (i.e., the same sample as in Table 2). The sample for columns (B) and (D) includes students from the same colleges and programs who dropped out in 2003–2009, and their earnings observations.

The dependent variable in columns (A) and (B) is an indicator for taking the exit exam. The regressions include program dummies and cohort dummies, where cohorts are defined by graduation year for college graduates and drop-out year for college drop-outs. We report the coefficient on the treatment variable δ_{pc}, which we define identically for graduation and drop-out cohorts.

The dependent variable in columns (C) and (D) is log average daily earnings. We report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. Column (C) is identical to column (A) in Table 3. The specification includes a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. Column (D) uses the same specification with cohorts and experience defined by drop-out year.

The means at the bottom of the table are calculated using only 2003–2004 graduates.

In all regressions, parentheses contain standard errors clustered at the program level.

Drop-outs are not a perfect counterfactual for graduates for several reasons. First, as shown in Table 5, drop-outs have smaller pre-treatment returns to reputation, although these returns are positive on average. Thus the information conveyed by drop-outs' college identities may differ from that of graduates. Further, drop-outs spent less time in college, and thus have less exposure to any potential accountability responses induced by the exit exam. However, if our main results were driven by accountability reforms rather than signaling, one would expect to see that drop-outs who stayed in college longer have treatment effects more similar to those of graduates. We find no evidence of this.

The drop-out placebo test is also consistent with balance regressions reported in Appendix B.G, which ask whether the exit exam rollout was correlated with changes in graduates' observable characteristics. If the field-specific introduction of the exit exams were correlated with trends in school or program choice, this should appear as changes in average reputation or Icfes scores across programs. There is little evidence of this channel. Changes in reputation and Icfes scores in programs with access to the exit exams are small and statistically insignificant.[98]

Appendix B.G also explores the effect of the exit exams on the probability of formal employment—a potential sample selection concern since we do not observe earnings for non-employed or informal workers. The estimated effect is not statistically significant and small relative to the mean formal employment rate.

In sum, the introduction of a new signal of skill—the field-specific college exit exams—reduced the return to reputation and increased the return to ability. These results are most consistent with an informational effect of the exit exams, and they provide evidence that college reputation signals individual ability to the labor market.

3.7. **Complementary effects of the exit exam.** There is suggestive evidence that the exit exam affected other outcomes. For example, Column (A) in Table 6 shows its impact on time to graduation. This estimate is from a standard differences in differences regression that includes program dummies, cohort dummies, and our treatment variable, δ_{pc}. The result suggests that individuals in programs with exam fields took about one quarter of a year longer to graduate. This is consistent with increased student effort, or with colleges taking steps to prepare students for the test. For instance, there is anecdotal evidence of colleges seeking to influence their students' performance, with activities ranging from "boot camp" preparation to more overt "gaming" via exclusion of certain students.[99]

[98] These results likely reflect high costs to switching programs in Colombia and the fact that our sample predominantly includes students who enrolled prior to the existence of any exit exams. Colombian colleges do not make it easy for students to change majors; switching essentially requires applying *de novo*.

[99] These results suggest that graduation cohort may be endogenous. We address this concern by estimating (8) with cohorts defined by *predicted* rather than actual graduation date, where predicted graduation is

Table 6. Complementary effects of the exit exam

	(A)	(B)	(C)
	Dependent variable		
	Years in college	Log daily earnings	Enrollees' Icfes scores
Exposed to exit exam (δ_{pc})	0.237	0.070	
	(0.110)	(0.019)	
Icfes reputation $\times$ $\delta_{p\tilde{c}}$			-0.162
			(0.053)
Exit exam reputation $\times$ $\delta_{p\tilde{c}}$			0.147
			(0.063)
N	146,052	581,802	485,350
R^2	0.132	0.201	0.277
# programs	39	39	39

Notes: The dependent variable in column (A) is graduation year minus enrollment year. The sample includes all students from Table 2. We report the coefficient on our treatment variable, δ_{pc}. The regression also includes program dummies and cohort dummies.

The dependent variable in column (B) is log average daily earnings for all observed experience levels (0–9 years). The sample includes all earnings observations from Table 2. In addition to δ_{pc}, the regression includes program dummies, cohort dummies, and a quadratic in experience interacted with program dummies.

The dependent variable in column (C) is individual Icfes percentile. The sample includes all students who enrolled in one of the 94 colleges and 39 problems in Table 2 between 2003 and 2009. We calculate Icfes and exit exam reputation using students who took the Icfes in 2000–2008, took the exit exam in 2009–2011 (when the exam was mandatory), and graduated from one the school-programs in our sample. We convert Icfes and exit exam scores into percentiles relative to this sample and within exit exam fields and years. We calculate reputation as means at the school-program level and normalize both measures so one unit represents ten percentile points in this distribution of exam takers. We define the treatment variable $\delta_{p\tilde{c}}$ using enrollment cohorts $\tilde{c}$, with $\delta_{p\tilde{c}} = \delta_{pc}$ for $\tilde{c} = c$. We report coefficients on the interactions of Icfes reputation and exit exam reputation with the treatment variable, $\delta_{p\tilde{c}}$. The regression includes dummies for program-cohort cells and interactions of both reputation measures with program dummies and cohort dummies.

In all regressions, parentheses contain standard errors clustered at the program level.

Using a similar specification, column (B) presents evidence that earnings increased by seven percent more in programs with early exam fields. This could have occurred if the exam improved match quality, raising overall productivity. It could also reflect students with access to the exam getting higher paying jobs at the expense college drop-outs and vocational school students, who are excluded from our sample.

Finally, we ask whether the exit exams altered individuals' school or program choices. This would be consistent with the government's stated intent. Column (C) explores how the

based on the year of enrollment. Appendix B.D shows that the results from this regression are similar to our benchmark specification; this suggests that selective graduation timing is not driving our main results.

ability of incoming students changed with the exit exam introduction. For this regression we define two measures of reputation using a population of graduates who took the exit exam in 2009–2011, when it was required of all graduates. We define Icfes reputation as mean Icfes percentile at the *school-program* level. Similarly, exit exam reputation is the school-program mean exit exam percentile. We convert Icfes and exit exam scores to percentiles within this population so that both reputation measures are on the same scale.

Icfes and exit exam reputations are highly correlated but not perfectly so. We suppose that the exit exam reputation contains new information, and that this information gradually became available to students entering college starting with the 2005 enrollment cohort.

Column (C) presents a specification analogous to the benchmark (8) with two key differences. First, the sample includes 2003–2009 *enrollees* rather than graduates, and we define students as treated by the exit exam ($\delta_{p\tilde{c}} = 1$) if they began a program p in an enrollment cohort $\tilde{c}$ after the introduction of the assigned field. Second, the dependent variable is the Icfes percentile of entering students, and we replace the independent variables R_s and τ_i with the school-program measures of Icfes and exit exam reputation. The reported coefficients in column (C) reflect how the correlations of Icfes and exit exam reputation with incoming students' Icfes scores changed with the exit exam rollout.[100]

The results show that in programs with exams, the ability of incoming students became more correlated with exit exam reputation, and less correlated with Icfes reputation. In other words, school-programs whose exit exam performance exceeded their average Icfes performance saw increases in the ability of their incoming classes. This suggests students selected different programs and/or colleges as new information on their quality became available.

4. College reputation and earnings growth

The previous section showed that college reputation plays a signaling role. This section asks whether college reputation serves *only* to signal ability as measured by admission scores.

[100] The full specification, of which column (C) reports only the γ^{τ} and γ^{exit} coefficients, is:

$$\tau_{ip\tilde{c}} = d_{p\tilde{c}} + (\mu_p + \mu_{\tilde{c}} + \gamma^{\tau}\delta_{p\tilde{c}})[\text{Icfes reputation}]_{sip} + (\nu_p + \nu_{\tilde{c}} + \gamma^{exit}\delta_{p\tilde{c}})[\text{Exit exam reputation}]_{sip} + e_{ip\tilde{c}}.$$

To do so, it explores the predictions from Proposition 2 (Section 2) on how college reputation correlates with initial earnings and with earnings growth.

4.1. **Sample.** We follow Farber and Gibbons (1996) and Altonji and Pierret (2001) in studying individuals making their initial transition to the labor force. We restrict our sample to individuals who: i) graduated in 2008 or 2009 (this allows us to observe earnings in the year of graduation and the next three years), and ii) entered the labor market immediately upon graduation and remained during four consecutive years (i.e., they did not attend graduate school or leave the formal labor force).[101] The results are thus not attributable to movements into and out of the labor market.

4.2. **Empirical specifications and results.** Our basic specification is:

$$w_{it} = d_{c_i t} + r_0 R_{s_i} + r\left(R_{s_i} \times t\right) + a_0 \tau_i + a\left(\tau_i \times t\right) + e_{it}. \tag{10}$$

The dependent variable, w_{it}, is log daily earnings for student i measured at potential experience t, which as before is employment year minus graduation year; $d_{c_i t}$ are graduation cohort c_i by experience t cell dummies; college reputation, R_{s_i}, and Icfes score, τ_i, are as before; r_0 is the return to reputation in the year of graduation, and r is the average change in the return to reputation from an additional year of potential experience; a_0 is period-zero return to ability, and a is the average yearly change in this return.[102] We report only coefficients on reputation, Icfes, and their interactions with experience, where the latter two are estimated using earnings only up to three years after graduation, the maximum we can observe for our sample of 2008–2009 graduates.

In estimating (10), our goal is not to identify the causal effect of reputation or admission scores. Our interest is in how their returns change with worker experience—the r and a coefficients—and whether these changes match the predictions from our signaling model.

[101] Appendix B.H provides further details on the sample.

[102] Formally, we parametrize the experience-specific r_t (and a_t) coefficients in equation (2) as $r_t = r_0 + r \times t$.

Table 7 estimates (10) both excluding and including Icfes terms, which yields the unconditional return to reputation and the conditional returns to reputation and Icfes. This corresponds to regressions (4) and (2) from Section 2 and the various subparts of Proposition 2.[103] We discuss results from each of these regressions separately in the subsections below.

4.2.1. *Unconditional return to reputation.* Column (A) of Table 7 estimates equation (10) including reputation but not Icfes terms, such that the estimates represent the unconditional return to reputation, r^u. The period-zero estimate shows that a one point increase in college reputation is associated with a ten percent increase in daily earnings in the year of graduation ($r_0^u \approx 0.10$). Proposition 2 predicts that the unconditional return to reputation should not change with experience, implying a zero coefficient on the interaction of reputation and experience. This arises because initial wages fully incorporate information employers observe, including college reputation. Thus reputation cannot predict innovations in wages; this is identical to wages being a martingale in Farber and Gibbons (1996).

Column (A) strongly rejects this prediction; the return to reputation *increases* with experience. Taken at face value, the coefficient implies that the advantage of having gone to a college with a one point greater reputation increases by about 50 percent within the first four years of employment. This contrasts with the results in Farber and Gibbons (1996) and Altonji and Pierret (2001), who find no evidence of an increasing effect of years of schooling, another educational trait workers might use to signal ability.

The contrast between the reputation and years of schooling results can also be depicted using earnings-experience profiles. Mincer (1974) noted that the wage profiles of workers with different schooling levels are approximately parallel throughout the earnings lifecycle. Panel A of Figure 4 replicates this finding using 2008–2012 household survey data from Colombia.[104] It plots the mean log hourly real wage among workers with two schooling

[103] Proposition 2 also contains predictions for regressions that include Icfes but not reputation terms. Appendix B.I shows that the results match the predictions: the unconditional return to Icfes increases with experience. This is consistent with findings in Farber and Gibbons (1996) and Altonji and Pierret (2001).
[104] In Figure 4, we define potential labor market experience as $\min(\text{age} - \text{years of schooling} - 6, \text{age} - 17)$. This definition differs from the one we use elsewhere in the paper (earnings year minus graduation year)

Table 7. Returns to reputation and ability, and experience interactions

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)
Reputation	0.101	0.071	0.079	0.055
	(0.017)	(0.017)	(0.017)	(0.016)
Reputation × t	0.017	0.017	0.012	0.008
	(0.003)	(0.003)	(0.003)	(0.002)
Icfes		0.033	0.024	0.017
		(0.003)	(0.002)	(0.002)
Icfes × t			0.006	0.002
			(0.001)	(0.001)
N	83,492	83,492	83,492	83,492
R^2	0.179	0.189	0.190	0.306
# colleges	130	130	130	130
Extra controls				Y

Notes: The dependent variable is log average daily earnings. The sample includes students in column (D) of Appendix Table B8 and earnings in the four years after graduation. Columns (A)-(C) estimate equation (10) excluding and including Icfes terms. In addition to the reported variables, both regressions include dummies for cohort-experience cells.

Column (D) adds the following controls to column (C): age at graduation, a gender dummy, dummies for eight mother's education categories, dummies for missing age and mother's education values, college program dummies, and dummies for college municipalities. Each control is interacted with a quadratic in experience.

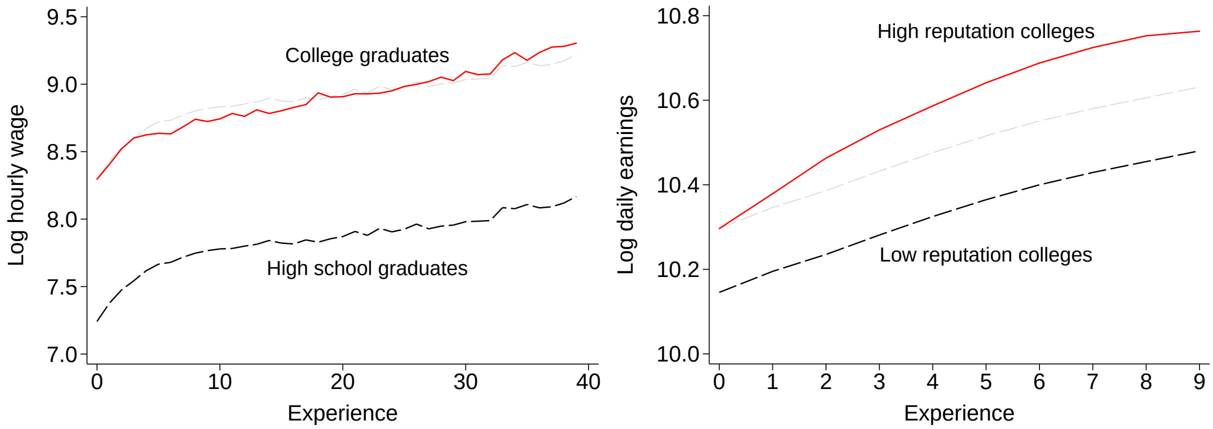
Parentheses contain standard errors clustered at the college level.

levels—completed high school and completed college—i.e., the gap between the two profiles is the college premium. This gap remains roughly constant across forty years of potential experience, consistent with results in the U.S. (Lemieux, 2006).[105]

Panel B uses our administrative data to plot earning profiles by college reputation. To match the cross-sectional analysis in Panel A, Panel B includes 2008–2012 earnings from all 2003–2012 college graduates. We plot mean log daily real earnings separately for graduates from high and low reputation colleges, defined by the median reputation. The earnings gap between the two profiles roughly doubles over the first ten years of experience, as indicated by the divergence of the high reputation profile from the light grey dashed line that is parallel to the low reputation profile.

because the Colombian household survey does not include school completion dates. However, the age and schooling definition matches those in Mincer's original analysis and in Altonji and Pierret (2001).

[105] The constant relationship between years of schooling and earnings in Colombia also holds in standard Mincerian regressions reported in Appendix B.J.



Panel A. Years of schooling

Panel B. College reputation

FIGURE 4. Earnings-experience profiles

Notes: Panel A includes high school and college graduates from the 2008–2012 monthly waves of the Colombia Integrated Household Survey (*Gran Encuesta Integrada de Hogares*). Lines depict the mean log hourly real wage (in 2008 pesos) for each schooling group, where we calculate means using survey weights. High school graduates are workers with exactly 11 years of schooling; college graduates have exactly 16 years of schooling. We define experience as $\min(\text{age} - \text{years of schooling} - 6, \text{age} - 17)$. The dashed light grey line is parallel to the high school profile starting from the college intercept.

Panel B includes 2003–2012 graduates from the 136 colleges represented in Figure 1 with earnings observations in 2008–2012. Lines depict the mean log daily real earnings (in 2008 pesos) for graduates from high and low reputation colleges, which we define by the unweighted median reputation of the 136 colleges. We define experience as $\text{age} - 16 - 6$ and omit levels of experience above nine years because they appear only for workers who took especially long to graduate. The dashed light grey line is parallel to the low reputation profile starting from the high reputation intercept.

These results thus suggest that the *slope* of workers' earnings-experience profiles increases with reputation. One potential explanation for this is that reputation may be imperfectly observed. Employers likely observe college identity, but they may not have access to our measure of reputation defined by mean Icfes scores. In this case employers would further learn about reputation through workers' output, resulting in a return to reputation that rises with experience. To address this possibility, we consider a stronger signaling prediction on regressions that also include individual admission scores.

4.2.2. *Conditional returns to reputation and ability.* Columns (B) and (C) of Table 7 add Icfes terms to the regression from column (A). We first add only Icfes scores, and then add an interaction term between Icfes and potential experience. Thus column (C) estimates equation (10) as written. In these joint specifications, the coefficients reflect the conditional returns to reputation and to ability from equation (2). As Proposition 2 predicts, the period-zero reputation coefficient in column (B) falls relative to its unconditional return in column (A). Consistent with employer learning about ability, column (C) also shows a positive and significant coefficient on the interaction of Icfes and experience.[106]

The main coefficient of interest is on the interaction of reputation with experience. Proposition 2 states that the conditional return to reputation should fall over time. This is similar to the Altonji and Pierret (2001) prediction for observable traits like race or schooling, but our definition of reputation yields a clean test of signaling. Since reputation is a group-level mean of Icfes, Icfes scores are a sufficient statistic for "admission exam" ability, α_i; conditional returns to reputation *mechanically* do not reflect the transmission of information on α_i. The conditional return to reputation should, therefore, decline with experience even if employers do not perfectly observe our measure of reputation; learning about reputation is reflected in the Icfes coefficients. Unlike Altonji and Pierret (2001), our model predicts a negative coefficient on Reputation$\times t$ even if there are interactions between ability, α_i, and human capital growth, h_{it}. These effects are also captured by the Icfes$\times t$ term.

In sum, if college reputation serves purely as a signal of ability, Proposition 2 predicts a negative coefficient on the interaction of reputation and experience. Column (C) clearly rejects this. The reputation-experience interaction, although smaller in magnitude than in column (A), is still positive and significant.

The increasing correlation of reputation and earnings is a descriptive result, but it is robust to a wide range of specifications and samples. For example, Column (D) of Table 7 adds controls for graduates' gender, age, socioeconomic status, college program, and regional market. All controls are interacted with a quadratic in potential experience to allow earnings

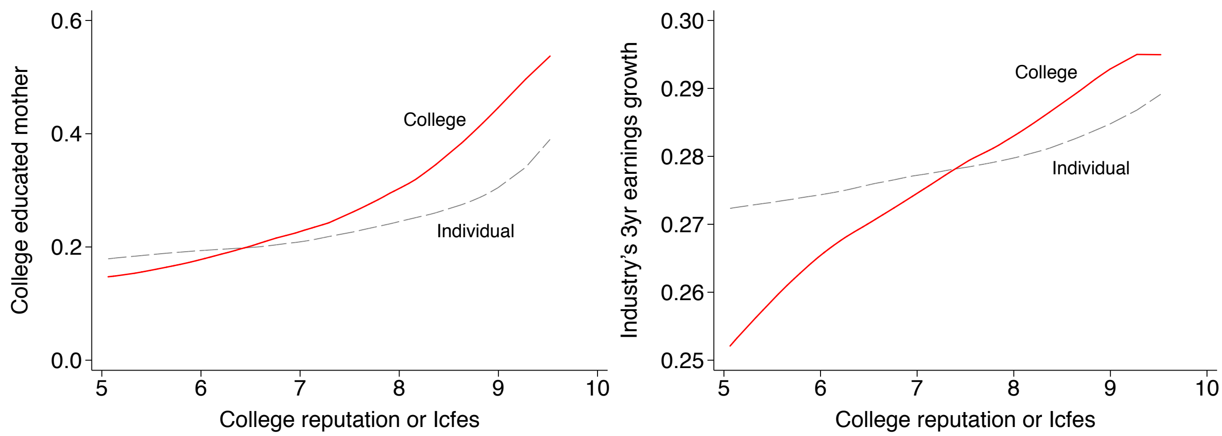
[106] The positive coefficient on the Icfes-experience interaction is similar to the Farber and Gibbons (1996) and Altonji and Pierret (2001) findings using Armed Forces Qualification Test (AFQT) scores as an unobserved characteristic. However, it is in contrast with findings in Arcidiacono, Bayer and Hizmo (2010), who also study AFQT scores but make a distinction between graduates who enter the labor market after high school and those who do so after college. For college graduates, they show that AFQT is strongly related to wages in the year of graduation, and this relationship changes little over the next ten years. Their conclusion is that AFQT revelation is complete for college graduates, and they suggest that this revelation occurs through college identity. Appendix B.I discusses one potential explanation for the difference in findings: sorting by ability in Colombia—although increasing—appears to be less extensive than in the U.S.

trajectories to vary with each characteristic. The coefficient on the reputation-experience interaction decreases slightly, but it is still highly significant and roughly of the same magnitude. Appendix B.K shows that this interaction term remains positive with further controls, different definitions of labor market experience, and in alternate samples. Appendix B.L also shows that the interaction term is positive when we define reputation at the college-program level rather than at the college level.

4.3. **Potential explanations for the increasing return to reputation.** The above results reject a model in which reputation relates to wages only as a signal of ability, α_i, and instead suggests that other attributes related to college membership influence earnings growth. In our model, these attributes are denoted by v_{s_i}, which we define to include both sorting on traits like socioeconomic status, and factors that contribute to skill acquisition at school such as teaching or peer effects. We suppose that employer expectations are given by $E\left\{v_{s_i}|R_{s_i}\right\} = v_0 + v_1 R_{s_i}$, where v_1 is the reputation premium. If v_1 is positive, an increasing return to reputation could arise for two reasons. First, if the market does not perfectly observe our measure of reputation, it may become increasingly correlated with wages as employers learn about other college membership attributes. Second, the return to reputation may rise if college membership attributes are related to human capital growth.

Figure 5 provides suggestive evidence that both of these channels may be at work. First, Panel A considers one potential component of v_{s_i}: socioeconomic status as measured by whether a student's mother has a college degree.[107] The x-axis contains reputation when observations are colleges, and Icfes when observations are individuals (the scale is the same). The solid line shows that as one moves from the college with the lowest reputation to that with the highest, the mean percentage of students with college-educated mothers increases from below 20 to above 50. The dashed line describes the individual-level relationship between students' Icfes scores and their mother's education, i.e., this is the relationship that would exist if sorting into colleges were by Icfes only. Socioeconomic sorting is less

[107] Similar patterns emerge for traits related to family income, parents' occupation, and geography.



Panel A. Socioeconomic status

Panel B. Industry's earnings growth

FIGURE 5. College membership attributes and their time-varying effects

Notes: The sample for Panel A is identical to Figure 1. The dependent variable is a dummy equal to one if a student's mother has a college/postgraduate degree.

The sample for Panel B includes any student in Panel A with a four-digit economic activity code from the Ministry of Social Protection. For each four-digit industry, we calculate the mean 2008 log daily earnings for 2005 college graduates and for 2008 college graduates. The dependent variable is the difference between the 2005 and 2008 cohort averages for the industry of each graduate's first job.

Dashed lines are local linear regressions of the dependent variable on Icfes percentile. Solid lines are local linear regressions of school means of the dependent variable on college reputation with weights equal to the number of graduates.

pronounced in this hypothetical scenario than in the actual one; i.e., there is more sorting across colleges on mother's schooling than is predicted by Icfes scores alone.[108] This is consistent with a positive reputation premium ($v_1 > 0$); sorting on mother's education is positively correlated with reputation. This could lead to a rising return to reputation if employers imperfectly observe both reputation and mother's education.

Second, Panel B shows that the reputation premium, v_1, may be correlated with human capital investment. The y-axis depicts the average three-year earnings growth in the industry of each graduate's first job. We define industries using four-digit codes, and we calculate earnings growth rates within industry as the mean difference in 2008 log earnings between 2005 and 2008 graduates. The dashed line shows the population-level relationship between industry earnings growth and Icfes scores. Graduates with 50^{th} percentile Icfes scores have

[108] The fact that Colombian financial aid markets are less developed suggests that straightforward ability to pay—beyond the lack of information or ability to take advantage of financial aid opportunities highlighted by Hoxby and Avery (2012) and Hoxby and Turner (2013*a*)—may account for some of the substantial role that socioeconomic status plays in college choice.

first jobs in industries where earnings increase by 27 percent within four years, and this growth rate rises by 1.5 percentage points across the Icfes distribution. The solid line shows that the relationship between earnings growth and college reputation is more pronounced. On average, graduates from colleges with reputations at the 50^{th} percentile enter industries in which earnings increase by only 25 percent within four years. Mean earnings growth is 4.5 percentage points higher in the industries that employ graduates from top colleges. In short, graduates from higher-ranked colleges obtain jobs in industries with greater earnings growth, and this relationship holds even for students with similar ability.

Table 8 further illustrates this point by displaying examples of these industries. For this table, we regress college reputation on individual Icfes scores and calculate the residuals. We display the top 10 and bottom 10 industries according to the average value of these residuals. This indicates whether graduates are sorting into industries beyond what their Icfes scores predict. For example, the top-ranked industry by this metric—securities trading—has a reputation residual of 0.52. This indicates that graduates whose first job is in securities trading come from colleges with 5.2 percentile points higher reputation than is predicted by their Icfes scores alone. Further, workers in securities trading experience rapid earnings growth, with earnings increasing by 67 percent within the first four years.

Many of the other industries that disproportionately employ graduates from top colleges are related to engineering, and they also tend to have high early-career earnings growth. By contrast, the mean earnings growth in the bottom 10 industries by reputation residual is 17 percentage points lower than that in the top 10. Many of these low-ranked industries are in the public sector, offering careers in government administration or elementary education.

These results suggest that the increasing return to reputation may reflect a career effect (Topel and Ward, 1992) in which better college reputation allows some individuals to be

Table 8. Industries of first employment ranked by reputation residuals

Top 10 industries

Industry	Mean reputation residual	3 year earnings growth
Securities trading	0.52	0.67
Higher education	0.41	0.27
Oil and gas extraction	0.31	0.50
Specialized animal breeding	0.30	0.15
Chemical manufacturing	0.28	0.52
Petroleum and natural gas extraction	0.28	0.47
Pharmaceutical products wholesale	0.27	0.42
Rubber products manufacturing	0.27	0.20
Cleaning products manufacturing	0.26	0.45
Radio and television	0.26	0.54
Top 10 average	0.33	0.36

Bottom 10 industries

Industry	Mean reputation residual	3 year earnings growth
Public transportation	-0.45	0.20
Financial cooperatives	-0.41	0.17
Social security services	-0.37	0.21
Preschool education	-0.34	0.22
Intermediate products wholesale	-0.33	0.19
Basic primary education	-0.32	0.19
Gambling services	-0.31	0.23
Public services and administration	-0.30	0.08
Elementary education facilities	-0.29	0.17
Soda and mineral water production	-0.27	0.32
Bottom 10 average	-0.34	0.19

Notes: This table includes industries with at least 100 graduates from the sample for Figure 5. We define industries and calculate earnings growth as in that figure. Reputation residuals are from a regression of college reputation on Icfes. We display only the top and bottom 10 industries ranked by the mean of these residuals. Averages for the top 10 and bottom 10 are weighted by the number of graduates in each industry.

matched to jobs with steeper wage profiles, or to firms that facilitate more on-the-job training. Higher reputation schools might also provide better networks (e.g., Kaufmann, Messner and Solis, 2013; Zimmerman, 2013) that ultimately make individuals more productive.[109]

[109] Other candidate explanations for the increasing return to reputation arise from violations of the assumptions of the competitive model itself. For example, labor contracts may be such that there is compression in starting wages. In U.S. law firms, for instance, it is not uncommon to observe entering associates being paid the same regardless of their law school of origin. Compensation may later diverge in a way correlated with an LSAT-based reputation measure (Heisz and Oreopoulos, 2002).

Our setting and data do not reveal whether the correlation between college reputation and earnings growth is due to unobserved dimensions of sorting or due to a causal effect of college identity. In particular, we cannot rule out the hypothesis that college reputation is merely serving as a *signal* for unobservable characteristics that themselves are related to human capital accumulation. Further, even if sorting into colleges occurs only on the ability dimension (α_i in our model), the increasing conditional return to reputation could arise because admission scores are imperfect measures of ability. Nonetheless, the widening of earnings profiles across Colombian colleges is starkly different from the parallel nature of earnings profiles across schooling levels. This may lead students to suspect that their choice of college quality matters for their earnings trajectories in a way that their choice of educational attainment might not.

5. Conclusion

Debates like those surrounding affirmative action suggest that college plays a key role in determining the distribution of opportunity. As a consequence a large literature studies the implications of college attendance. Some papers (e.g., Card, 1995) ask if college has a causal return, while others (e.g., Goldin and Katz, 2008*a*) consider the evolution and determinants of the college wage premium.

Such work does not address the dilemma faced by the millions of students who—having decided to go to college—must choose one. The size of the test preparation industry, for example, suggests that students and parents believe that college choice is important, and that life opportunities are better if one goes to a better college. We call the process by which students are matched to colleges and subsequently to jobs, "the big sort."

This paper has explored the role that college reputation plays in the big sort. Specifically, we have shown that if colleges are selective and more able students choose more "reputable" colleges, then one can produce a single dimensional measure of college *reputation*. We chose a particular measure—the average admission test score of a college's graduates—because it allows a clean examination of signaling mechanisms.

We showed that, consistent with work on other markets, employers use college reputation to make inferences about individual graduates. Specifically, while the cross-sectional data are consistent with this, we exploited a natural experiment in Colombia to show that providing more information about student skill reduces the importance of reputation. Thus college identity performs a signaling function, and students may be right to worry about *which* college in addition to *whether* college. In other words, we find support for MacLeod and Urquiola's (2015) assumption that labor markets do not immediately observe all individual characteristics (such as Icfes or AFQT scores), and college membership may transmit some of them.

However, we also find that signaling is not the whole story. Even after controlling for admission scores, a graduate's starting earnings and earnings *growth* are positively correlated with her college's reputation. These results are consistent with the hypothesis that colleges add to skill, and that their value added varies systematically with their reputation. Although we cannot establish that this is a causal link, these correlations matter because they are observable—students may notice that individuals from better schools seem to get careers with higher earnings trajectories, which may lead them to prefer more reputable schools.

The purpose of the big sort is to match individuals to jobs. A literature documents significant differences in compensation across firms and occupations that cannot be explained by worker ability (Krueger and Summers, 1988; Gibbons and Katz, 1991; Abowd et al., 1999). Our evidence is consistent with the hypothesis that college choice is partially driven by what students *believe* is the consequence of choice. We find evidence that in Colombia, as in the United States, students prefer colleges that are more selective; this in turn leads employers to offer higher wages to the graduates of such colleges. These results illustrate that increasing access to college will not necessarily reduce wage and income inequality. The big sort is a complex system that moves moves individuals from high school to their first job. Our finding that the exit exams reduced the reputation premium suggests other countries could use analogous policies to address wage inequality.

A. Theoretical Appendix

This appendix presents a complete version of the theory in Section 2, which incorporates college reputation into the literature on information and wage formation (Jovanovic, 1979; Farber and Gibbons, 1996; Altonji and Pierret, 2001). We define a measure of reputation, specify a model of wage setting, and conclude with derivations of propositions that are the basis for our empirical analyses in Sections 3 and 4.

A.1. **Ability, admission scores, and college reputation.** We let α_i denote the log ability of student i, where we use the term ability to represent the type of aptitude measured by pre-college admission tests. We suppose $\alpha_i \sim N(0, \frac{1}{\rho^\alpha})$, where $\rho^\alpha = \frac{1}{\sigma_\alpha^2}$ is the precision of α_i. For simplicity we assume all variables are mean zero and normally distributed, and we characterize their variability using precisions.

We define two measures of α_i. First, we observe each student's score on a college admission exam. We denote it by τ_i and assume it provides a noisy measure of ability:

$$\tau_i = \alpha_i + \epsilon_i^\tau,$$

where ρ^τ is the precision of τ_i. Second, we define the *reputation* of a college s to be the mean admission score of its graduates, and denote it by R_s:

$$R_s = E\left\{\tau_i | i \in s\right\} = \frac{1}{n_s} \sum_{i \in s} \tau_i,$$

where n_s is the number of graduates from college s. Note that this definition implies that for student i randomly selected from college s_i, we can view reputation as a signal of the individual admission score and write:

$$R_{s_i} = \tau_i + \epsilon_i^{R,\tau}, \tag{A1}$$

where $\rho^{R,\tau}$ is the precision of $\epsilon_i^{R,\tau}$. We define college reputation in this way because it provides a clear benchmark against which to test various hypotheses on how reputation relates to wages. Since reputation is a noisy measure of the admission score, then τ_i is a sufficient statistic for reputation in the following sense:

$$E\{\alpha_i|\tau_i, R_{s_i}\} = E\{\alpha_i|\tau_i\}. \tag{A2}$$

If colleges were perfectly selective, then all students at school s would have the same admission score, such that $\rho^{R,\tau} = \infty$. In practice colleges are never perfectly selective; hence we can suppose that our measure of reputation is less precise than admission scores: $\rho^{R,\tau} < \infty$.

Given (A1) we can write:

$$R_{s_i} = \alpha_i + \epsilon_i^{\tau} + \epsilon_i^{R,\tau},$$

and let $\rho^R < \rho^\tau$ be the precision of the error term $\epsilon_i^{\tau} + \epsilon_i^{R,\tau}$. Given these definitions for the signals of student ability, we use Bayes' rule to derive three structural parameters that depend on the precisions of ability, admission scores, and reputation:[110]

$$E\{\alpha_i|\tau_i\} = \frac{\rho^\tau}{\rho^\alpha + \rho^\tau}\tau_i = \pi^{\alpha|\tau}\tau_i \tag{A3}$$

$$E\{\alpha_i|R_{s_i}\} = \frac{\rho^R}{\rho^\alpha + \rho^R}R_{s_i} = \pi^{\alpha|R}R_{s_i} \tag{A4}$$

$$E\{R_{s_i}|\tau_i\} = \frac{\rho^{R,\tau}}{\rho^\tau + \rho^{R,\tau}}\tau_i = \pi^{R|\tau}\tau_i. \tag{A5}$$

Since $0 < \rho^R < \rho^\tau < 1$, the first two parameters satisfy $0 < \pi^{\alpha|R} < \pi^{\alpha|\tau} < 1$. The extent to which colleges are selective is given by $\pi^{R|\tau} \in [0, 1]$, where $\pi^{R|\tau} = 0$ if students are randomly allocated to colleges, and $\pi^{R|\tau} = 1$ if students perfectly sort by admission scores. Since the number of colleges is less than the number of students, the assumption of normally distributed ability and test scores is sufficient to ensure $\pi^{R|\tau} < 1$.

A.2. **Employers' information and wage setting process.** We let θ_i denote the log skill of student i and suppose it is given by:

$$\theta_i = \alpha_i + v_{s_i}.$$

[110] Notice that, for example, $E\{\alpha_i|\tau_i\} = \frac{\rho^\tau}{\rho^\alpha+\rho^\tau}\tau_i + \frac{\rho^\alpha}{\rho^\alpha+\rho^\tau}E\{\alpha_i\}$, but we have set $E\{\alpha_i\} = 0$.

Skill includes both pre-college ability, α_i, and v_{s_i}, which we will interpret as attributes related to an individual's membership at college s_i. These can include factors that contribute to skill formation at school, such as teaching or peer effects, as well as access to alumni networks. They can also include individual traits (not perfectly correlated with α_i) along which individuals select into colleges, such as family income or individual motivation.

We suppose that the market sets log wages, w_{it}, equal to expected skill given the information, I_{it}, available regarding worker i in period t:

$$w_{it} = E\left\{\theta_i | I_{it}\right\} + h_{it}.$$

h_{it} is time-varying human capital growth due to experience and on the job training; it may also vary with graduation cohort and other time-invariant control variables. We follow the literature on the Mincer wage equation (see Lemieux, 2006) and net out human capital growth to consider equations of the form:

$$\hat{w}_{it} = w_{it} - h_{it} = E\left\{\theta_i | I_{it}\right\}.$$

We use log wages net of human capital growth, $\hat{w}_{it}$, to focus on the time-invariant component of skill that is generated by schooling and revealed over time. Farber and Gibbons (1996) observe that this leads to a martingale representation for wages. In particular, it implies that for $t \geq 1$, innovations in wages cannot be forecasted with current information:

$$E\left\{\hat{w}_{it} - \hat{w}_{i,t-1} | I_{i,t-1}\right\} = 0.$$

We suppose that employers' information set, I_{it}, includes college reputation, R_{s_i}.[111] While employers likely care about individuals' pre-college ability as captured by R_{s_i}, they also care about other attributes related to graduates' post-college skill. We therefore define a college's

[111] Employers likely observe college identity, but they may not perfectly observe our measure of reputation. Below we discuss how our definition helps to address the possibility that this assumption does not hold.

labor market reputation as the expected skill of its graduates: $\mathscr{R}_s = E\{\theta_i | i \in s\}$. It follows that $\theta_{i\in s} \sim N(\mathscr{R}_{s_i}, \frac{1}{\rho^{\mathscr{R}}})$, where $\rho^{\mathscr{R}}$ denotes the precision of $\mathscr{R}_s$.[112]

Our data do not contain $\mathscr{R}_s$, and it may differ from R_s if colleges with higher reputation provide more value added or select students students based upon dimensions of ability that are not observable to us. For instance, if colleges prefer motivated students, and students prefer more value added, there will be a positive correlation between our measure of reputation, R_s, and other college membership attributes, v_s. To allow for this possibility we suppose v_s satisfies $E\{v_s|R_s\} = v_0 + v_1 R_s$, where $v_1 > 0$ is the reputation premium.

Thus, employers observe a signal of worker i's skill given by the labor market reputation of her college of origin:

$$\begin{aligned} \mathscr{R}_{s_i} &= E\{\alpha_i + v_{s_i} | R_{s_i}\} \\ &= \pi^{\alpha|R} R_{s_i} + v_0 + v_1 R_{s_i}. \end{aligned} \tag{A6}$$

In other words, labor market reputation captures employers' expectations of ability, α_i, and attributes related to college membership, v_s, under the assumption that they observe our measure of reputation, R_s.

Following Farber and Gibbons (1996), firms observe other signals of worker skill—not including labor market reputation—that are available at the time of hiring but are not visible to us. For instance, employers might obtain such information by conducting job interviews or obtaining references. We denote this information by:

$$y_i = \alpha_i + v_s + \epsilon_i, \tag{A7}$$

with associated precision ρ^y. Importantly, we assume y_i does not include τ_i; that is, employers do not observe a graduate's individual admission test score. This is consistent with the assumption in the employer learning literature that AFQT scores are unobserved.

[112] The precision, $\rho^{\mathscr{R}}$, could also be indexed by s and hence be school-specific. We did not find robust evidence that the variance has a clear effect on earnings, and so set this aside for further research.

Lastly, employers observe signals related to worker output after employment begins:

$$y_{it} = \alpha_i + v_s + \epsilon_{it}, \tag{A8}$$

where ϵ_{it} includes human capital growth and other fluctuations in worker output. We suppose these are observed *after* setting wages in each period t, where $t = 0$ stands for the year of college graduation. We let $\bar{y}_{it} = \frac{1}{t+1}\sum_{k=0}^{t} y_{ik}$ denote mean worker output and suppose that the precision of y_{it} is time invariant and denoted by $\rho^{\bar{y}}$.[113]

The market's information set regarding student i in period t is thus $I_{it} = \{\mathscr{R}_{s_i}, y_i, y_{i0}, ..., y_{i,t-1}\}$. Bayesian learning implies that log wages net of human capital growth satisfy:

$$\hat{w}_{it} = \pi_t^{\mathscr{R}} \mathscr{R}_{s_i} + \pi_t^{y} y_i + \left(1 - \pi_t^{\mathscr{R}} - \pi_t^{y}\right) \bar{y}_{i,t-1}, \tag{A9}$$

where the weights on the signals are given by:

$$\begin{aligned} \pi_t^{\mathscr{R}} &= \frac{\rho^{\mathscr{R}}}{\rho^{\mathscr{R}} + \rho^{y} + t\rho^{\bar{y}}} \\ \pi_t^{y} &= \frac{\rho^{y}}{\rho^{\mathscr{R}} + \rho^{y} + t\rho^{\bar{y}}}. \end{aligned} \tag{A10}$$

Note that $\pi_t^{\mathscr{R}}, \pi_t^{y} \to 0$ as wages incorporate the new information from worker output.

A.3. **Regressions on characteristics in our data.** Equation (A9) describes employers' wage setting process given the information they observe, I_{it}. We do not observe I_{it}, and instead derive the implications of the wage equation for regressions on characteristics in our data. We use regressions that include controls for experience and graduation cohort to capture the time-varying effects (recall from above that $\hat{w}_{it} = w_{it} - h_{it}$). Here we focus upon the implications of the model for the relationship between the signals of individual ability

[113] The assumption that the precision of y_{it} is time stationary also follows Farber and Gibbons (1996). We note that this assumption implies that any human capital growth included in ϵ_{it} is not serially correlated.

and wages net of human capital growth. In particular, we consider three regressions:

$$\text{(A11)} \qquad \hat{w}_{it} = r_t^u R_{s_i} + e_{it}^R$$

$$\text{(A12)} \qquad \hat{w}_{it} = a_t^u \tau_i + e_{it}^\tau$$

$$\text{(A13)} \qquad \hat{w}_{it} = r_t R_{s_i} + a_t \tau_i + e_{it},$$

where the e_{it} variables are residuals. We define the coefficient on reputation in (A11), r_t^u, as the unconditional *return to reputation* at time t. The coefficient on the admission score in (A12), a_t^u, is the unconditional *return to ability.* Specification (A13) estimates the conditional return to reputation, r_t, and the conditional return to ability, a_t.

To derive the values of these coefficients, we plug the definitions for $\mathscr{R}_s$, y_i, and $\bar{y}_{i,t-1}$ from (A6)-(A8) into the wage equation (A9):

$$\begin{aligned} \hat{w}_{it} &= \pi_t^{\mathscr{R}} \left(\pi^{\alpha|R} R_{s_i} + v_0 + v_1 R_{s_i}\right) + \pi_t^y \left(\alpha_i + v_s + \epsilon_i\right) \\ &\quad + \left(1 - \pi_t^{\mathscr{R}} - \pi_t^y\right)\left(\alpha_i + v_s + \bar{\epsilon}_{i,t-1}\right) \\ \text{(A14)} \qquad &= v_0 + v_1 R_{s_i} + \pi_t^{\mathscr{R}} \pi^{\alpha|R} R_{s_i} + \left(1 - \pi_t^{\mathscr{R}}\right) \alpha_i + \epsilon_{it}^w, \end{aligned}$$

where $\epsilon_{it}^w = \left(1 - \pi_t^{\mathscr{R}}\right)\left(v_s - v_0 - v_1 R_{s_i} + \bar{\epsilon}_{i,t-1}\right) + \pi_t^y \left(\epsilon_i - \bar{\epsilon}_{i,t-1}\right).$

To generate predictions for our three regressions, we take expectations of (A14) with respect to reputation, R_s, and the admission score, τ_i. For this we use the structural parameters defined by (A3)-(A5). Regression (A11) is given by:

$$\begin{aligned} E\left\{\hat{w}_{it} | R_{s_i}\right\} &= v_0 + v_1 R_{s_i} + \pi_t^{\mathscr{R}} \pi^{\alpha|R} R_{s_i} + \left(1 - \pi_t^{\mathscr{R}}\right) \pi^{\alpha|R} R_{s_i} \\ \text{(A15)} \qquad &= v_0 + \left(v_1 + \pi^{\alpha|R}\right) R_{s_i}. \end{aligned}$$

Regression (A12) is given by:

$$\begin{aligned} E\left\{\hat{w}_{it} | \tau_i\right\} &= v_0 + v_1 \pi^{R|\tau} \tau_i + \pi_t^{\mathscr{R}} \pi^{\alpha|R} \pi^{R|\tau} \tau_i + \left(1 - \pi_t^{\mathscr{R}}\right) \pi^{\alpha|\tau} \tau_i \\ \text{(A16)} \qquad &= v_0 + \left(v_1 \pi^{R|\tau} + \pi^{\alpha|\tau} - \pi_t^{\mathscr{R}} (\pi^{\alpha|\tau} - \pi^{\alpha|R} \pi^{R|\tau})\right) \tau_i. \end{aligned}$$

Finally, regression (A13) requires taking expectations of (A14) with respect to both R_{s_i} and τ_i, and it uses the sufficient statistic assumption (A2):

$$
\begin{aligned}
E\left\{\hat{w}_{it}|R_{s_i},\tau_i\right\} &= v_0 + v_1 R_{s_i} + \pi_t^{\mathscr{R}}\pi^{\alpha|R}R_{s_i} + \left(1-\pi_t^{\mathscr{R}}\right)\pi^{\alpha|\tau}\tau_i \\
&= v_0 + \left(v_1 + \pi_t^{\mathscr{R}}\pi^{\alpha|R}\right)R_{s_i} + \left(\pi^{\alpha|\tau} - \pi_t^{\mathscr{R}}\pi^{\alpha|\tau}\right)\tau_i. \qquad \text{(A17)}
\end{aligned}
$$

From equations (A15)-(A17) we can define the coefficients on reputation and the admission score in the regressions (A11)-(A13):

$$r_t^u = v_1 + \pi^{\alpha|R} \qquad \text{(A18)}$$

$$a_t^u = v_1\pi^{R|\tau} + \pi^{\alpha|\tau} - \pi_t^{\mathscr{R}}\left(\pi^{\alpha|\tau} - \pi^{\alpha|R}\pi^{R|\tau}\right) \qquad \text{(A19)}$$

$$r_t = v_1 + \pi_t^{\mathscr{R}}\pi^{\alpha|R} \qquad \text{(A20)}$$

$$a_t = \pi^{\alpha|\tau} - \pi_t^{\mathscr{R}}\pi^{\alpha|\tau}. \qquad \text{(A21)}$$

These coefficients form the basis for Propositions 1 and 2.

A.4. **Predictions for the introduction of a college exit exam.** In Section 3 we ask how the conditional returns to reputation and ability were affected by the introduction of another measure that graduates could use to signal their ability—a college exit exam. We suppose that the exit exam increases the amount of information regarding the skill of student i contained in y_i, such that its precision is $\rho^{y,exit} > \rho^y$ when the exit exam is offered. This could originate in multiple channels, including students listing exit exam scores on their CVs, receiving reference letters as a result of their performance, or modifying job search behavior after learning their position in the national distribution of exam takers.

From the definition of $\pi_t^{\mathscr{R}}$ in (A10), note that $\rho^{y,exit} > \rho^y$ implies $\pi_t^{\mathscr{R},exit} < \pi_t^{\mathscr{R}}$ for every t, where $\pi_t^{\mathscr{R},exit}$ is the weight on labor market reputation in the presence of the exit exam. Let $\delta_i = 1$ if and only if a student is exposed to the possibility of writing the exit exam. We can rewrite the joint regression (A13) as follows:

$$
\begin{aligned}
\hat{w}_{it} &= (1-\delta_i)\left(r_t R_{s_i} + a_t \tau_i\right) + \delta_i \left(r_t^{exit} R_{s_i} + a_t^{exit} \tau_i\right) + e_{it}^{exit} \\
&= \left(r_t R_{s_i} + a_t \tau_i\right) + \delta_i \left(\beta_t^r R_{s_i} + \beta_t^a \tau_i\right) + e_{it}^{exit}, \qquad \text{(A22)}
\end{aligned}
$$

where:

$$
\begin{aligned}
\beta_t^r &= r_t^{exit} - r_t \\
&= \left(\pi_t^{\mathscr{R},exit} - \pi_t^{\mathscr{R}}\right) \pi^{\alpha|R} < 0, \qquad \text{(A23)}
\end{aligned}
$$

$$
\begin{aligned}
\beta_t^a &= a_t^{exit} - a_t \\
&= \left(\pi_t^{\mathscr{R}} - \pi_t^{\mathscr{R},exit}\right) \pi^{\alpha|\tau} > 0. \qquad \text{(A24)}
\end{aligned}
$$

The simplifications of β_t^r and β_t^a follow from the values of the conditional returns to reputation and ability in (A20) and (A21).[114] This in turn implies:

Proposition 1. If wages are set to expected skill given the available information (equation (A9)), then the introduction of an exit exam reduces the return to college reputation ($\beta_t^r < 0$) and increases the return to ability ($\beta_t^a > 0$).

We examine the empirical evidence related to Proposition 1 in Section 3.

A.5. **Predictions for earnings growth.** In Section 4, we describe how the returns to reputation and ability change with experience, t, thereby comparing college reputation to other signals of ability studied in the literature. The coefficient values given by equations (A18)-(A21) imply the following proposition:

Proposition 2. If wages are set equal to expected skill given the available information (equation (A9)), then:

[114] Since our regressions use log wages, the experience profiles reflect the reduction in uncertainty as information accumulates about the worker. Experience profiles can therefore differ for individuals with $d_i = 1$ and $\delta_i = 0$. To account for such effects, in the regressions below we include controls for experience that vary with individuals' potential access to the exit exams.

(1) The unconditional return to reputation, r_t^u, does not change with experience.

(2) The unconditional return to ability, a_t^u, rises with experience.

(3) The conditional return to reputation, r_t, is smaller than the unconditional return, and with experience falls to v_1, the reputation premium.

(4) The conditional return to ability, a_t, is smaller than the unconditional return, and rises with experience.

Part (1) holds because r_t^u does not depend on t. Part (2) holds because $-\pi_t^{\mathscr{R}}$ is increasing with t, $\pi^{\alpha|\tau} > \pi^{\alpha|R}$, and $\pi^{R|\tau} < 1$. Part (3) follows from $\pi_t^{\mathscr{R}}$ decreasing with t, $\pi_t^{\mathscr{R}} < 1$, and $\pi^{\alpha|R} > 0$. Part (4) holds if $v_1, \pi^{\alpha|\tau}, \pi^{\alpha|R}, \pi^{R|\tau}, \pi_t^{\mathscr{R}} > 0$.

Note that if reputation is imperfectly observed, its unconditional return should rise with experience, mirroring the prediction for admission scores in part (2). The possibility that employers do not perfectly observe reputation does not alter the prediction in part (3), however, as any employer learning about reputation should be reflected in the conditional admission score coefficients.

We take the predictions from Proposition 2 to the data in Section 4 and Appendix B.9.

A.6. **Signaling vs. accountability effects of the exit exam.** In this section we develop three additional propositions that we discuss but do not present formally in the paper. This provides a way of testing for effects of the exit exams other than those related to signaling. For example, the exams may have prompted colleges to undertake accountability-related reforms, such as modifying their curricula or adding test-preparation sessions. Individuals may also have worked harder in preparation for the exams.

Such accountability effects would affect the skills a student developed while in college, rather than their pre-college ability. In our model, such post-enrollment attributes are given by v_s, which satisfies $E\{v_s|R_s\} = v_0 + v_1 R_s$. Thus to test how the exit exams affected the return to reputation, we allow the v_1 term to differ between students with and without access to the exams. Specifically, we let v_1^{exit} denote college membership attributes for students with access to exit exams, while v_1 represents such traits for students without access to exams. With this extra notation we can derive predicted effects on the conditional returns to reputation and ability using equations (A20), (A21), and (A22):

$$\begin{aligned} \beta_t^r &= r_t^{exit} - r_t \\ &= \left(v_1^{exit} - v_1\right) + \left(\pi_t^{\mathscr{R},exit} - \pi_t^{\mathscr{R}}\right)\pi^{\alpha|R}, \end{aligned} \tag{A25}$$

$$\begin{aligned} \beta_t^a &= a_t^{exit} - a_t \\ &= \left(\pi_t^{\mathscr{R}} - \pi_t^{\mathscr{R},exit}\right)\pi^{\alpha|\tau}. \end{aligned} \tag{A26}$$

Note that the reputation effect of the exit exams, β_t^r, has an extra term $(v_1^{exit} - v_1)$ relative to that in (A23), but the ability effect, β_t^a, is identical. This arises because reputation, R_s, is a better predictor of college membership attributes than individual admission scores, τ_i.

Now suppose that the introduction of the exit exams had accountability effects but no implications for signaling based on college reputation. In terms of the model, this means that $v_1^{exit} \neq v_1$ but $\pi_t^{\mathscr{R},exit} = \pi_t^{\mathscr{R}}$. Equations (A25) and (A26) thus yield a non-zero effect on the conditional return to reputation, β_t^r, and a zero effect on the conditional return to ability, β_t^a. This result is summarized in the follow proposition:

Proposition 3. If the introduction of an exit exam has accountability effects ($v_1^{exit} \neq v_1$) but no signaling effects ($\pi_t^{\mathscr{R},exit} = \pi_t^{\mathscr{R}}$), then the conditional return to college reputation should change ($\beta_t^r \neq 0$) and the conditional return to ability should be unaffected ($\beta_t^a = 0$).

It is also useful to explore the effects of the exit exams on the *unconditional* returns to reputation and ability. Using equations (A18) and (A19), we have:

$$
\begin{aligned}
\beta_t^{u,r} &= r_t^{u,exit} - r_t^u \\
&= v_1^{exit} - v_1, \qquad \text{(A27)}
\end{aligned}
$$

$$
\begin{aligned}
\beta_t^{u,a} &= a_t^{u,exit} - a_t^u \\
&= \left(v_1^{exit} - v_1\right)\pi^{R|\tau} + \left(\pi_t^{\mathscr{R}} - \pi_t^{\mathscr{R},exit}\right)\left(\pi^{\alpha|\tau} - \pi^{\alpha|R}\pi^{R|\tau}\right). \qquad \text{(A28)}
\end{aligned}
$$

If we assume that the exit exams had signaling effects ($\pi_t^{\mathscr{R},exit} < \pi_t^{\mathscr{R}}$) but no accountability effects ($v_1^{exit} = v_1$), then we should observe $\beta_t^{u,r} = 0$ and $\beta_t^{u,a} > 0$. Note also that under these assumptions the effect of the exit exams on the unconditional return to ability in (A28) should be smaller than that on the conditional return to ability in (A26). This is summarized in the following proposition:

Proposition 4. If the introduction of an exit exam has signaling effects ($\pi_t^{\mathscr{R},exit} < \pi_t^{\mathscr{R}}$) but no accountability effects ($v_1^{exit} = v_1$), then the unconditional return to college reputation should not change ($\beta_t^{u,r} = 0$) and the unconditional return to ability should increase but be smaller than the conditional return ($0 < \beta_t^{u,a} < \beta_t^a$).

If instead we assume that the introduction of the exit exams had accountability effects ($v_1^{exit} \neq v_1$) but no signaling effects ($\pi_t^{\mathscr{R},exit} = \pi_t^{\mathscr{R}}$), then we should find a non-zero effect on both the unconditional return to reputation, $\beta_t^{u,r}$, and to ability, $\beta_t^{u,a}$. Importantly, these effects should have the same sign, as the changes in v_1 can be measured by either R_s or τ_i when we include these terms individually. This yields the proposition:

Proposition 5. If the introduction of an exit exam has accountability effects ($v_1^{exit} \neq v_1$) but no signaling effects ($\pi_t^{\mathscr{R},exit} = \pi_t^{\mathscr{R}}$), then the unconditional returns to reputation and ability should change ($\beta_t^{u,r} \neq 0$, $\beta_t^{u,a} \neq 0$) and should have the same sign.

Propositions 3, 4, and 5 provide a rich set of predictions that allow us to explore whether the exit exam effects are likely to be the result of signaling or accountability mechanisms. We discuss the empirical evidence related to these predictions in Section 3.6.4.

B. Empirical Appendix

This appendix provides details on the samples and further robustness checks for the empirical analyses in Sections 3 and 4.

B.1. **Matching college programs to exit exam fields.** This section describes the matching of exit exam fields to college programs, which allows us to define a treatment variable for Section 3. Columns (A) and (B) in Table B1 (and the table notes) list the 55 field exams that were introduced between 2004 and 2007. In 2009, a "generic competency" (*competencias genéricas*) exam was made available for programs without a corresponding field.

Although the exit exams were field-specific, during the period we study there was no formal system assigning college majors to exam fields. This match is necessary to determine which majors were treated. We therefore perform this assignment ourselves, using three different approaches. In our benchmark approach, we consider all college majors belonging to the Ministry of Education's 54 core knowledge groups. These groups—which we label programs—aggregate approximately 2,000 college major names that vary across and within schools. For instance, the Ministry might combine a major named Business Administration at one college with one labeled Business Management at another if it considers that these have similar content. We assign each of the 54 programs to one of the 55 exam fields if one of the key words in the program name appears in the name of the field exam. We assign programs without any matching key words to the generic competency exam introduced in

Table B1. Exit exam fields, college programs, and sample selection

(A) Year	(B) Exit exam field	(C) College program	(D) Program area	(E) Graduates	(F) Colleges	(G) Included
2004	Medicina veterinaria	Medicina veterinaria	Agronomy	2,055	2	Y
	Zootecnia	Zootecnia	Agronomy	1,144	1	
	Ingeniería agronómica y agronomía	Agronomía	Agronomy	84		
	Administración	Administración	Business	28,406	46	Y
	Contaduría	Contaduría pública	Business	15,712	36	Y
	Economía	Economía	Business	8,646	21	Y
	Licenciatura exams (seven in total)	Educación	Education	16,910	21	Y
	Ingeniería industrial	Ingeniería industrial y afines	Engineering	12,331	25	Y
	Ingeniería de sistemas	Ingeniería de sistemas, telematica y afines	Engineering	11,312	25	Y
	Ingeniería civil	Ingeniería civil y afines	Engineering	7,347	19	Y
	Ingeniería electrónica	Ingeniería electrónica, telecomunicaciones y afines	Engineering	7,385	14	Y
	Arquitectura	Arquitectura y afines	Engineering	4,400	12	Y
	Ingeniería mecánica	Ingeniería mecánica y afines	Engineering	4,639	9	Y
	Ingeniería ambiental	Ingeniería ambiental, sanitaria y afines	Engineering	3,804	8	Y
	Ingeniería de alimentos	Ingeniería agroindustrial, alimentos y afines	Engineering	1,443	5	Y
	Ingeniería química	Ingeniería química y afines	Engineering	3,439	4	Y
	Ingeniería eléctrica	Ingeniería eléctrica y afines	Engineering	1,490	3	Y
	Ingeniería agronómica y agronomía	Ingeniería agronómica, pecuaria y afines	Engineering	1,474	3	Y
	Ingeniería agrícola	Ingeniería agrícola, forestal y afines	Engineering	903	1	
	Enfermería	Enfermería	Health	7,927	19	Y
	Medicina	Medicina	Health	7,767	8	Y
	Fisioterapia	Terapias	Health	5,126	8	Y
	Odontología	Odontología	Health	2,616	7	Y
	Bacteriología	Bacteriología	Health	2,211	6	Y
	Nutrición y dietética	Nutrición y dietética	Health	1,019	3	Y
	Optometría	Optometría, otros programas de ciencias de la salud	Health	629	3	Y
	Psicología	Psicología	Social sciences	11,726	24	Y
	Derecho	Derecho y afines	Social sciences	15,934	21	Y
	Comunicación e información	Comunicación social, periodismo y afines	Social sciences	6,441	16	Y
	Trabajo social	Sociología, trabajo social y afines	Social sciences	4,201	7	Y
2005	Biología	Biología, microbiología y afines	Natural sciences	3,257	5	Y
	Química	Química y afines	Natural sciences	1,712	1	
	Matemática	Matemática, estadística y afines	Natural sciences	551	1	
	Física	Física	Natural sciences	396	1	
	Geología	Geología, otros programas de ciencias naturales	Natural sciences	379		
2006	Instrumentación quirúrgica	Instrumentación quirúrgica	Health	1,416	5	Y
2007	Educación física, recreación, deportes y afines	Deportes, educación física y recreación	Social sciences	405		
2009	Competencias genéricas	Ingeniería administrativa y afines	Engineering	2,225	5	Y
		Ingeniería de minas, metalurgia y afines	Engineering	1,554	2	Y
		Otras ingenierías	Engineering	720	2	Y
		Ingeniería biomédica y afines	Engineering	358	1	
		Diseño	Fine arts	4,609	7	Y
		Publicidad y afines	Fine arts	1,320	5	Y
		Artes plásticas, visuales y afines	Fine arts	2234	4	Y
		Música	Fine arts	462		
		Artes representativas	Fine arts	55		
		Otros programas asociados a bellas artes	Fine arts	15		
		Salud pública	Health	225	1	
		Ciencia política, relaciones internacionales	Social sciences	2,641	4	Y
		Lenguas modernas, literatura, lingüística y afines	Social sciences	841	4	Y
		Antropología, artes liberales	Social sciences	668	3	Y
		Geografía, historia	Social sciences	647	2	Y
		Bibliotecología, otros de ciencias sociales y humanas	Social sciences	97	1	
		Filosofía, teología y afines	Social sciences	548		

Notes: Columns (A) and (B) list exit exam fields and their year of introduction. *Licenciatura* includes seven exams covering pedagogical training intended for school teachers of preschool education, natural sciences, social sciences, humanities, math, French, and English. Column (C) shows the Ministry of Education's 54 core knowledge groups that we call programs. We match exam fields to programs using the method described in footnote 115. Thirteen fields did not match any program: 2004) Fonoaudiología, medicina veterinaria y zootecnia, terapia ocupacional; 2005) Ingeniería agroindustrial, ingeniería forestal, ingeniería de petróleos, técnico en electrónica y afines, técnico en sistemas y afines, tecnológico en electrónica y afines, tecnológico en sistemas y afines; 2006) Normalistas superiores, técnico profesional en administración y afines, tecnología en administración y afines. Column (D) shows eight program "areas" the Ministry of Education uses to categorize these 54 programs. Column (E) lists the number of 2003–2009 graduates with non-missing Icfes scores that appear in the earnings records. Column (F) reports the number of colleges offering each program after trimming and balancing the sample. A "Y" in column (G) indicates programs included in our final sample. See Appendix B.3 for details on trimming, balancing, and selecting programs.

2009.[115] Column (C) in Table B1 shows the resulting match of programs and exit exam fields. This is a more detailed version of the match displayed in Table 1 of Section 3.

A second approach is to match programs to fields based on the most common exam students in each program took in 2009, when all fields and the generic exam were available. In this alternate procedure, we compute the percentage of 2009 test takers in each program that took a field exam introduced in 2004, 2005, 2006, or 2007, and the percentage that took the generic exam. We assign each program to an exit exam year using the maximum of these five percentages. This procedures differs from the name-matching method in only four programs: mathematics (*matemáticas, estadística y afines*), chemistry (*química y afines*), agricultural and forest engineering (*ingeniería agrícola, forestal y afines*), and mining and metallurgical engineering (*ingeniería de minas, metalurgia y afines*).[116] Mining and metallurgical engineering is the only one of these four programs that appears in our final sample.

In 2011, the agency that administers the exit exam began assigning programs from each college to one of 17 "reference groups," and they required each group to take different components. For a third procedure for matching programs to fields, we obtained these reference groups for the 2013 exam, but this test is significantly different from the 2004–2009 tests covered in Table B1; it contains numerous subject-specific modules and several common components. We assume reference groups that took the generic exam module in 2013 had no exit exam field for the 2003–2009 cohorts. We assume all other reference groups received an

[115] We define a "key word" as one that appears in only one of the 54 program names, ignoring articles and removing plural endings. If a program has no key word because its name is duplicated in other programs, we set the key word to the entire program name, ignoring the words "and related" ("*y afines*"). If we match a program to multiple fields, we use the field with an identical name if possible or the field with the earliest introduction date otherwise. In the Ministry of Education's classification, *educación* is the program group for all education degree (*licenciatura*) programs, so we assign *educación* to the seven *licenciatura* exams introduced in 2004 and exclude these exams for matching with other programs.

[116] This procedure matches mathematics and chemistry to the generic exam rather than the mathematics and chemistry fields because the exit exam fields were less widely adopted in these programs. Agricultural and forest engineering is assigned to the 2005 exam group rather than the agricultural engineering field because 2009 test takers most commonly took the forest engineering field exam. Lastly, mining and metallurgical engineering is assigned to the 2005 exam group rather than the generic exam because students most commonly took the petroleum engineering field (*ingeniería de petróleos*).

Table B2. Sensitivity of exit exam effects to field-program matching

Dependent variable: log average daily earnings

	(A)	(B)	(C)
	Benchmark specification	Most frequent 2009 field	Reference groups for 2013 exam
Reputation × δ_{pc}	−0.041	−0.040	−0.026
	(0.017)	(0.017)	(0.020)
Icfes × δ_{pc}	0.017	0.017	0.014
	(0.006)	(0.006)	(0.004)
N	581,802	581,802	681,077
R^2	0.258	0.258	0.234
# programs	39	39	17

Notes: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. All regressions include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. The sample for each regression includes experience 0–9. Parentheses contain standard errors clustered at the program level.

Column (A) is identical to column (A) in Table 3. All other columns estimate this same specification with different definitions of the treatment variable δ_{pc}. Column (B) defines treatment using the most common exam field taken by students from each program in 2009. Column (C) defines treatment using the Colombian Institute for Educational Evaluation's 2013 "reference groups." See Appendix B.1 for details.

exit exam field starting with the 2005 cohort except for the natural sciences group, which received an exit exam field starting with the 2006 cohort. We then select the sample following all procedures described in Appendix B.3 with reference groups as our program variable.

B.2. **Sensitivity of exit exam effects to field-program matching.** Table B2 tests the sensitivity of our exit exam results to the three field-program matching procedures described in Appendix B.1. In column (A), we replicate our benchmark results from Table 3, which matches exit exam fields to college programs based on their names.

In column (B), we match fields to programs based on the most common exam students in each program took in 2009. In our final sample, the assignment of programs to exit exam fields under this procedure differs from that in the name-matching method for only one program. The estimated effects in column (B) are thus similar to our benchmark specification. We use the name-matching procedure for our main results, however, because students' exam choices are potentially endogenous.

Column (C) uses the exam agency's 17 "reference groups" as our definition of programs. Our results are qualitatively similar under this procedure, though the reputation effect is smaller in magnitude with this coarser definition of treatment. We prefer using the Ministry of Education's programs to define treatment because they align better with the granularity of the 2004–2009 exam fields.

B.3. **Section 3 sample.** In this section we describe how we select the cohorts, programs, and colleges we include in our empirical analysis in Section 3.

Our sample includes the 2003–2009 graduation cohorts. While our dataset covers students who enrolled in 1998–2012, there are few graduates before 2003 because students typically take at least five years to graduate. Further, we drop the 2010–2012 graduates in order to focus cleanly on the period in which signals of field-specific skill were introduced into a subset of fields. This is no longer clearly the case after the 2009 cohort due to several structural changes in the exit exams.[117]

We define potential labor market experience, t, as calendar year minus graduation cohort, and drop any earnings observations prior to graduation. Our final sample therefore includes 2008–2012 earnings for 2003–2008 graduates and 2009–2012 earnings for 2009 graduates.

Two factors motivate how we select programs and colleges for our sample. First, our empirical specification will estimate the return to reputation for students in the same program and cohort. This return is imprecisely estimated when there are few students in the same school, program, and cohort, or when few colleges offer a given program. Second, our identification comes from the staggered introduction of the exit exam fields. Columns (D) and (E) in Table B1 show the Ministry of Education's categorization of programs into eight "program areas," and the number of 2003–2009 graduates in each program. Exam fields for most large programs in business, health, and engineering were introduced immediately in 2004. Field exams were delayed or never created for mostly smaller programs in natural

[117] In 2009 common components in English and reading comprehension were introduced for all test takers, and a required generic exam for those not taking a field test was made available. Furthermore, 22 of the field exams were removed in 2010–2011 and replaced with more aggregate exam modules.

sciences, social sciences, and fine arts. Identification thus directly counteracts precision by requiring we include smaller programs offered by fewer colleges.

Our final sample balances these considerations. We begin with 367,526 graduates from 133 colleges.[118] Roughly 25 percent of these students never appear in our earnings records, and about 20 percent are missing Icfes scores or program variables. Excluding these leaves 225,856 graduates.

We then calculate the number of earnings observations across all experience levels for each school-program-cohort and drop cells below the 10^{th} percentile number of observations. After trimming, we drop school-programs with missing cohorts to balance the composition across all seven cohorts. Trimming eliminates ten percent of the sample with non-missing data and balancing the sample eliminates about 25 percent more. After this step, there are 147,788 graduates from 94 colleges.

Finally, in order to identify a return to college reputation, each program must be offered by at least two colleges. Column (F) in Table B1 shows the number of colleges that offer each program after trimming and balancing the sample. We exclude any program offered at a single school.

The final sample includes the 39 programs with a "Y" in column (G) and any colleges that offer those programs after trimming and balancing. This covers 146,052 graduates from 94 colleges. We observe four years of earnings per student on average, resulting in 581,802 total observations.

Table 2 in Section 3 displays summary statistics for the final sample. Table B3 here displays analogous summary statistics for students excluded in the process described above. The excluded population is about 50 percent larger in size than the sample for Section 3, but it has fewer total earnings observations. In general excluded students have only slightly lower Icfes scores but attend colleges with reputations that are on average four percentile

[118] As stated we consider only graduating students who obtained 4–5 year degrees, the equivalent of bachelor degrees in the U.S. The sample for Section 3 begins with 136 colleges, but three of these only have 2010–2011 graduates in our records.

Table B3. Summary statistics for Section 3 excluded students

Variable	2004	2005	2006	2007	2009	All
	Year program received exit exam					
# graduates in 2003–2009	183,206	7,042	1,240	622	29,364	221,474
# earnings obs. in 2008–2012	440,635	18,648	2,747	1,808	74,090	537,928
# programs	30	5	1	1	18	55
# colleges	133	29	10	6	86	133
Reputation	6.97	8.25	6.33	6.59	7.63	7.09
	(1.21)	(1.08)	(0.87)	(0.66)	(1.11)	(1.23)
Icfes	7.52	9.03	6.18	6.20	7.80	7.61
	(2.39)	(1.32)	(2.45)	(2.34)	(2.19)	(2.35)
Log average daily earnings	10.83	10.96	10.62	10.33	10.76	10.82
	(0.67)	(0.72)	(0.57)	(0.45)	(0.71)	(0.68)
Return to reputation	0.080	0.040	0.060	1.393	0.041	0.075
	(0.021)	(0.055)	(0.033)	(0.121)	(0.032)	(0.017)
Return to ability	0.020	0.022	-0.020	-0.013	0.065	0.028
	(0.005)	(0.029)	(0.012)	(0.027)	(0.015)	(0.005)

Notes: This table presents summary statistics for 2003–2009 graduates in our records that are excluded from the main analysis sample in Section 3 (i.e., those not included in Table 2). All variables are defined identically as in Table 2. Note that one reason we excluded these students is due to missing values on certain variables, so the statistics in this table are averages for only students who have values of each variable.

points lower. Their average return to reputation is about six percentage points lower, but they have a similar average return to Icfes.[119]

B.4. **Sensitivity of exit exam effects to sample selection.** Table B4 tests the sensitivity of our exit exam results to the sample selection procedure described in Appendix B.3. Column (A) of this table reprints our benchmark results from column (A) of Table 3.

In our benchmark sample, we calculate the number of observations in each school-program-cohort cell and exclude cells below the 10^{th} percentile. We exclude small school-program-cohorts because our empirical specification requires that we calculate returns to reputation and Icfes within each program and cohort, and these returns are imprecisely estimated with few observations. After trimming, we balance the panel so that our sample includes only school-programs that appear in all seven cohorts (2003–2009).

[119] In most cases, sample sizes are large enough that we can reject equality of mean characteristics between included (Table 2) and excluded (Table B3) students.

Table B4. Sensitivity of exit exam effects to sample selection

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
		School-program-cohort trimming			# colleges in each program		
	Benchmark specification	No trimming	5^{th} percentile	25^{th} percentile	3 or more colleges	4 or more colleges	Predicted cohorts
Reputation × δ_{pc}	−0.041	−0.035	−0.040	−0.038	−0.042	−0.036	−0.044
	(0.017)	(0.015)	(0.016)	(0.031)	(0.018)	(0.017)	(0.018)
Icfes × δ_{pc}	0.017	0.006	0.012	0.020	0.016	0.015	0.017
	(0.006)	(0.007)	(0.008)	(0.009)	(0.006)	(0.006)	(0.010)
N	581,802	671,840	618,489	452,080	575,321	563,752	650,015
R^2	0.258	0.256	0.260	0.254	0.248	0.247	0.241
# programs	39	48	41	31	35	30	39
Trim percentile	10^{th}	0^{th}	5^{th}	25^{th}	10^{th}	10^{th}	10^{th}
Colleges/program	2+	2+	2+	2+	3+	4+	2+
Grad. cohorts	Actual	Actual	Actual	Actual	Actual	Actual	Predicted

Notes: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. All regressions include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. The sample for each regression includes experience 0–9. Parentheses contain standard errors clustered at the program level.

Column (A) is identical to column (A) in Table 3. All other columns estimate this same specification using different samples.

Columns (B)-(D) use different percentiles for the number of observations below which we drop small school-program-cohort cells. Our main specification in column (A) trims school-program-cohort cells below the 10^{th} percentile in terms of number of observations. Columns (B), (C), and (D) use no trimming, the 5^{th} percentile, and the 25^{th} percentile. In all cases we balance the sample after trimming so that each remaining school-program appears in all seven cohorts in our sample. All other sample selection methods follow as described in Appendix B.3.

Columns (E) and (F) use different minimums for the number of schools that we require to offer each program. Our main specification in column (A) requires the bare minimum necessary to identify a return to reputation within each program: each program must be offered by two or more colleges. Columns (E) and (F) require that each program must be offered by three or more, and four or more, colleges. All other sample selection methods follow as described in Appendix B.3.

Column (G) addresses the possible endogeneity of graduation cohort discussed in footnote 99. We create a new sample based on the year students entered college, $\tilde{c}$, rather than the year they graduated, c. Most university programs in Colombia have an official duration of ten semesters, so we define predicted graduation date as $\tilde{c}+5$. We include only students whom we predict to graduate in 2003–2009. In other words, this sample covers graduates who enrolled in 1998–2004, regardless of when they graduated. Because selective graduation also affects labor market experience, we replace our measure of potential experience with years since expected graduation, $\tilde{t} = y - (\tilde{c}+5)$, where y is calendar year. We modify our benchmark specification (8) by replacing graduation cohort, c, with enrollment cohort, $\tilde{c}$, and potential experience, t, with predicted potential experience, $\tilde{t}$. We define the treatment variable $\delta_{p,\tilde{c}+5}$ as before with expected rather than actual graduation year—i.e., $\delta_{p,\tilde{c}+5} = \delta_{pc}$ with $c = \tilde{c}+5$.

Columns (B)-(D) use different percentiles for the number of observations below which we drop small school-program-cohort cells. Columns (B), (C), and (D) use no trimming, the 5^{th} percentile, and the 25^{th} percentile. In all cases we balance the sample after trimming so that each remaining school-program appears in all seven cohorts. All other sample selection methods follow as in Appendix B.3. The signs are consistent across all trimming thresholds, though the reputation coefficient loses significance when we trim at the 25^{th} percentile, and the Icfes coefficient loses significance when we trim at the 5^{th} percentile or do not trim. The variation in statistical significance across trimming thresholds reflects the data demands of our empirical strategy, though the consistency of the signs is reassuring.

Columns (E) and (F) use different minimums for the number of schools that we require to offer each program. Our main specification in column (A) requires the bare minimum necessary to identify a return to reputation within each program: each program must be offered by two or more colleges. Columns (E) and (F) require that each program must be offered by three or more, and four or more, colleges. All other sample selection methods follow as in Appendix B.3. Our results are not sensitive to this choice.

Table 6 in Section 3 shows that the exit exam may have increased time to graduation. This suggests that graduation cohort may be endogenous in the estimation of our reputation and Icfes effects. Column (G) addresses this issue by defining a sample based on *predicted* graduation cohort rather than actual graduation cohort. Most university programs in Colombia have an official duration of ten semesters, so we define predicted graduation as five years after enrollment. The sample includes students predicted to graduate in 2003–2009—i.e., those who enrolled in 1998–2004—regardless of when they actually graduated. Because selective graduation also affects labor market experience, we redefine potential experience as years since predicted graduation, rather than years since actual graduation. The specification for column (G) is otherwise identical to column (A) with cohort and potential experience defined by predicted graduation.

Column (G) shows that the estimates from this regression are similar to our benchmark specification, which suggests that selective graduation timing is not driving our main results.

B.5. **Returns to reputation and ability by program-cohort.** Our regression analysis in Section 3 is derived from a two-step estimation procedure. The first step equation (6) estimates conditional returns to reputation and ability separately for each program and cohort. The second step equation (7) relates these returns to the availability of the exit exam, captured in our treatment variable, δ_{pc}. Our benchmark specification (8) combines these two steps into a single regression.

To illustrate this procedure, Table B5 presents program-cohort specific returns from a regression similar to the first-step specification (6). Columns (A)-(C) display the 39 programs in our sample and the introduction year of the exit exam field we assigned to each program (see Table B1). Columns (F) and (G) present the conditional returns to reputation for each program and cohort, $\hat{r}_{pc}$, except we use only two cohort groups: students who graduated before the introduction of any exit exams (2003–2004) and those who graduated after the first field exams became available (2005–2009). Column (H) reports the difference between pre- and post-exam returns for each program. Columns (I)-(K) similarly show the program-cohort returns to ability, $\hat{a}_{pc}$, and their difference.

As shown in Table 2, most of our identification comes from a comparison of programs that received exit exams in the first year ("2004 programs") and programs that never received an exam during our period of analysis ("2009 programs"). We can thus illustrate our main results with a simple 2×2 difference in differences analysis using these two groups. The bottom rows of Table B5 show the average pre- and post-exam returns to reputation and ability for 2004 and 2009 programs.[120] The boxed numbers report the 2×2 difference in differences estimates. For example, the return to reputation declined from 13.8 to 9.8 percent in 2004 programs, but was unchanged at 3.0 percent in 2009 programs. The difference in differences estimate is thus roughly -4 percent, similar to our benchmark coefficient in Table 3. The 2×2 estimate for the return to ability is 2.1 percent, which is also close to our benchmark result.

[120] Averages are weighted by each coefficient's inverse squared standard error from the first-step regression.

Table B5. Returns to reputation and ability by program and cohort

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)
Exam Year	Program area	Program	N	Colleges	Return to reputation 2003–04	Return to reputation 2005–09	Return to reputation Diff.	Return to ability 2003–04	Return to ability 2005–09	Return to ability Diff.
2004	Agronomy	Medicina veterinaria	1,808	2	-0.09	-0.04	0.05	0.03	-0.02	-0.05
	Business	Administración	85,325	46	0.18	0.14	-0.05	0.04	0.03	-0.01
	Business	Contaduría pública	49,714	36	0.18	0.09	-0.09	0.03	0.03	0.00
	Business	Economía	25,879	21	0.21	0.11	-0.10	0.05	0.03	-0.02
	Education	Educación	40,195	21	0.10	0.05	-0.05	0.02	0.02	0.00
	Engineering	Ingeniería industrial y afines	41,309	25	0.23	0.15	-0.08	0.04	0.02	-0.02
	Engineering	Ingeniería de sistemas, telematica y afines	28,526	25	0.19	0.15	-0.04	0.05	0.04	-0.01
	Engineering	Ingeniería civil y afines	24,334	19	0.10	0.09	-0.01	0.02	0.01	-0.01
	Engineering	Ingeniería electrónica, telecom. y afines	15,657	14	0.19	0.13	-0.06	0.08	0.02	-0.06
	Engineering	Arquitectura y afines	11,701	12	0.07	0.05	-0.03	0.02	0.01	-0.01
	Engineering	Ingeniería mecánica y afines	9,659	9	0.27	0.19	-0.08	-0.02	0.01	0.03
	Engineering	Ingeniería ambiental, sanitaria y afines	7,251	8	0.14	0.06	-0.08	0.03	0.01	-0.02
	Engineering	Ingeniería agroindustrial, alimentos y afines	2,889	5	0.03	0.12	0.09	0.08	0.03	-0.05
	Engineering	Ingeniería química y afines	7,630	4	0.44	0.25	-0.19	0.02	0.06	0.04
	Engineering	Ingeniería eléctrica y afines	2,320	3	0.13	0.00	-0.13	0.01	0.03	0.01
	Engineering	Ingeniería agronómica, pecuaria y afines	1,559	3	0.30	0.26	-0.04	0.09	0.01	-0.07
	Health	Enfermería	27,824	19	0.06	0.08	0.02	0.01	0.01	0.00
	Health	Medicina	13,520	8	0.01	0.00	-0.01	0.03	0.01	-0.02
	Health	Terapias	12,211	8	0.05	0.03	-0.03	0.10	0.06	-0.04
	Health	Odontología	5,211	7	0.02	0.01	0.00	-0.03	0.00	0.02
	Health	Bacteriología	6,304	6	0.03	0.05	0.02	0.03	0.02	-0.01
	Health	Nutrición y dietética	3,635	3	-0.20	-0.22	-0.03	-0.03	-0.01	0.02
	Health	Optometría, otros prog. de ciencias de la salud	1,895	3	0.08	-0.01	-0.09	-0.01	0.00	0.02
	Social sciences	Psicología	35,506	24	0.11	0.07	-0. 04	0.03	0.02	-0.01
	Social sciences	Derecho y afines	39,608	21	0.12	0.10	-0.02	0.02	0.01	-0.01
	Social sciences	Comunicación social, periodismo y afines	19,523	16	0.18	0.13	-0.05	0.02	0.02	0.00
	Social sciences	Sociología, trabajo social y afines	7,442	7	0.08	0.05	-0.03	0.02	0.01	-0.01
2005	Natural sciences	Biología, microbiología y afines	7,418	5	0.04	0.17	0.13	0.01	-0.02	-0.02
2006	Health	Instrumentación quirúrgica	4,516	5	-0.22	0.03	0.26	0.02	0.02	0.01
2009	Engineering	Ingeniería administrativa y afines	3,936	5	0.16	0.12	-0.04	0.07	0.03	-0.04
	Engineering	Ingeniería de minas, metalurgia y afines	2,367	2	-0.01	-0.16	-0.15	0.13	0.02	-0.11
	Engineering	Otras ingenierías	1,558	2	0.11	0.13	0.02	-0.08	-0.02	0.05
	Fine arts	Diseño	12,641	7	0.02	0.04	0.02	0.04	0.01	-0.02
	Fine arts	Publicidad y afines	3,412	5	-0.01	0.00	0.02	0.03	0.02	-0.01
	Fine arts	Artes plásticas, visuales y afines	6,704	4	-0.19	-0.15	0.03	0.05	0.03	-0.03
	Social sciences	Ciencia política, relaciones internacionales	4,806	4	0.04	0.09	0.05	0.04	0.01	-0.03
	Social sciences	Lenguas modernas, literatura, ling. y afines	3,101	4	0.18	0.13	-0.05	0.10	0.03	-0.07
	Social sciences	Antropología, artes liberales	2,160	3	-0.22	0.04	0.26	0.04	-0.03	-0.07
	Social sciences	Geografía, historia	748	2	-1.84	21.05	22.89	0.25	0.00	-0.25
		2004 programs	528,435	94	0.138	0.098	-0.041	0.030	0.021	-0.009
		2009 programs	41,433	21	0.030	0.030	0.000	0.048	0.018	-0.030
		Difference			0.109	0.068	-0.041	-0.018	0.003	0.021

Notes: Column (A) lists the introduction year of the exit exam field assigned to each of the 39 programs in our sample, which appear in column (C). Column (B) contains the program area of each program. Column (D) shows the number of earnings observations in our sample, and column (E) shows the number of colleges in our sample offering each program. See Table B1 and Appendices B.1 and B.3 for details.

Columns (F) and (G) report conditional returns to reputation for each program from specification (6) using only two cohort groups: 2003–2004 and 2005–2009. In other words, the returns to reputation coefficients are from a regression of log average daily earnings on interactions of reputation and Icfes with dummies for cells defined by programs and the 2003–2004 and 2005–2009 cohort groups. This regression includes an experience quadratic interacted with program dummies and dummies for program-cohort cells. Column (H) displays the difference between columns (F) and (G). Columns (I) and (J) report conditional returns to ability from the same specification, and column (K) displays their difference.

Averages at the bottom are weighted by each coefficient's inverse squared standard errors from this regression.

Table B5 also helps to explain the estimates in columns (E) and (F) of Table 3. These estimates restrict identification to programs with similar pre-exit exam returns to reputation and ability. Columns (F) and (I) in Table B5 show these pre-exam returns.[121] Though 2004 programs generally have higher returns to reputation and lower returns to ability, there are exceptions to both cases. This allows us to match 2004 programs to delayed exit exam programs that have similar returns.

B.6. **Exit exam effects on the returns to other characteristics.** An alternative hypothesis for our main results is that the exit exams affected signaling on observable characteristics other than college reputation. To explore this hypothesis, in Table B6 we replicate our benchmark regression (column (A) in Table 3) replacing the reputation terms with other individual characteristics that may be at least partially observable to employers.

Column (A) replicates our benchmark results using college reputation. Columns (B)-(D) replace reputation with indicators for gender, mother's education, and family income, respectively. In each regression, the return to these other characteristics declines with the exit exam, but none of the effects are statistically significant. Further, the Icfes effects in all these regressions are also statistically insignificant. In column (E), we include terms for all characteristics jointly; only the Icfes and reputation effects are statistically significant.

Although we cannot rule out signaling effects on characteristics not included in our data, the results in Table B6 provide evidence that the strongest effects of the exit exams were on the returns to college reputation.

B.7. **Balance tests.** Section 3.6.6 discusses three balance tests that ask if the exit exam rollout was correlated with sorting into colleges or programs, or with the probability of formal employment. Table B7 shows the results from these balance tests. These estimates are from simple differences in differences regressions that include program dummies, cohort

[121] In actuality, the pre-exit exam returns in Table B5 are estimated in a regression that also includes 2005–2009 graduates, while the pre-exit exam returns used for columns (E)-(F) of Table 3 are from a specification including only 2003–2004 cohorts. This has little effect on the returns displayed in Table B5.

Table B6. Exit exam effects on the returns to other characteristics

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)
Icfes × δ_{pc}	0.017	0.006	0.007	0.007	0.018
	(0.006)	(0.006)	(0.006)	(0.008)	(0.006)
Reputation × δ_{pc}	−0.041				−0.036
	(0.017)				(0.016)
Male × δ_{pc}		−0.021			−0.023
		(0.015)			(0.015)
College mother × δ_{pc}			−0.036		−0.026
			(0.035)		(0.036)
High income × δ_{pc}				−0.039	−0.031
				(0.030)	(0.038)
N	581,802	581,645	576,945	576,332	574,803
R^2	0.258	0.232	0.236	0.237	0.263
# programs	39	39	39	39	39
Mean return to ability	0.029	0.068	0.062	0.064	0.024
Mean return to reputation	0.133				0.125
Mean return to gender		0.038			0.038
Mean return to mother's ed			0.123		0.076
Mean return to income				0.115	0.066

Notes: All regressions are identical to the benchmark specification in column (A) of Table 3, but they substitute the reputation terms in this regression with other characteristics. All columns report coefficients on the interactions of these characteristics with the treatment variable δ_{pc}. Regressions include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of Icfes and other characteristics with program and cohort dummies. Parentheses contain standard errors clustered at the program level.

College mother is an indicator for a student's mother having a technical college or university degree. High income is an indicator for a student's family income being greater than 300 percent of the minimum wage.

The mean returns at the bottom of the table are calculated using only 2003–2004 graduates.

dummies, and our indicator for exposure to the exit exams, δ_{pc}. The dependent variable for each regression is listed in the column header.

In columns (A) and (B), the dependent variables are college reputation, R_s, and Icfes percentile, τ_i. If the field-specific introduction of the exit exam was correlated with trends in school or program choice, this should appear as changes in average reputation or Icfes scores across programs. There is little evidence of this channel. Reputation increased by only 0.3 percentile points more in programs with access to the exit exams, while Icfes scores increased by 0.7 percentile points relative to programs without exam fields. Neither effect is statistically significant.

Table B7. Balance tests

	(A)	(B)	(C)
	Dependent variable		
	Reputation	Icfes	Has formal earnings
Exposed to exit exam (δ_{pc})	0.026	0.070	0.017
	(0.051)	(0.078)	(0.016)
N	146,052	146,052	890,809
R^2	0.204	0.146	0.044
# programs	39	39	39
Dependent variable mean	7.44	7.84	0.66

Notes: All columns report coefficients on the treatment variable δ_{pc}. Parentheses contain standard errors clustered at the program level.

The dependent variables in columns (A) and (B) are reputation and Icfes. The sample includes all students from Table 2. Each regression includes program dummies and cohort dummies.

The dependent variable in column (C) is an indicator for appearing in our earnings records at each year in 2008–2012. We include multiple observations per student for any level of potential labor market experience in 0–9 years. The sample includes all students from Table 2 plus graduates from the same programs and colleges who never appear in the earnings records. The regression includes program dummies, cohort dummies, and a quadratic in experience interacted with program dummies.

Column (C) expands our main sample to include students and years for which we do not observe earnings. The dependent variable is a dummy equal to one if the graduate appears in our earnings records t years after graduation.[122] The mean of this variable is 66 percent, and the remaining 34 percent is a composite measure of unemployment, informal employment, non-participation in the labor market, and pursuit of further education. The estimate suggests that formal employment increased 1.7 percentage points more in programs with exit exam fields, but this effect is not statistically significant. The small magnitude of this coefficient mitigates the concern that our main treatment effects are driven by sample selection in terms of who appears in the formal labor market.

B.8. **Section 4 sample.** In Section 4, we follow Farber and Gibbons (1996) and Altonji and Pierret (2001) in studying a sample of individuals making their initial transition to the long-term labor force. This subsection describes the construction of this sample.

[122] This regression also includes a quadratic in experience interacted with program dummies to control for program-specific time effects on the likelihood of formal employment.

The columns of Table B8 divide 2008–2009 graduates according to their post-college labor market paths. We choose these cohorts because our earnings records cover 2008–2012, which allows us to observe earnings in the year of graduation and the next three years.

Column (A) includes any student who enrolled in a specialization, masters, or doctorate program by 2011, the last year for which we have graduate education records. Columns (B)-(D) categorize those who did not enter graduate school by the number of years for which they have formal earnings in the first four years after graduation.[123] Column (B) includes students who never appear in our earnings records, while column (D) contains students who have formal earnings in each of the first four years. Column (C) contains students who move into and out of the formal labor force—those with 1–3 years of earnings.

Column (A) shows that 16 percent of 2008–2009 college graduates attend graduate school. These students tend to be from more reputable colleges, and they have higher Icfes scores and more educated mothers. Column (D) shows that 28 percent of students enter the formal labor force for four consecutive years after graduation. These students are typically of higher ability than graduates who do not transition to the long-term labor market, and they are are slightly more likely to be male.[124]

Our sample for Section 4 includes only students in column (D). Our estimates are therefore from a population with higher ability, but importantly, they are not attributable to movements into and out of the labor force; all results come from earnings changes within the formal labor market.

B.9. **Unconditional return to ability.** This section presents results related to the Proposition 2 predictions for the unconditional return to ability (Icfes).[125]

[123] We consider workers as having formal earnings if they have at least one monthly earnings observation in a given year.

[124] F-tests for each characteristic strongly reject the hypothesis of joint equality across the four columns.

[125] We note that the Icfes percentiles we use in Section 4 are conceptually similar to those in Section 3, but they are based on different data sources. In Section 4, we compute Icfes percentiles using data from the Colombian Institute for Educational Evaluation (see the notes to Figure 1). This yields a relatively continuous variable. In Section 3, we use Icfes percentiles from the Ministry of Education records because the data from the Colombian Institute for Educational Evaluation do not cover our earliest graduating cohorts. The Ministry of Education computes Icfes percentiles in a similar manner (i.e., position relative to

Table B8. Transition from college to the labor market

2008–2009 college graduates

Variable	(A) Went to graduate school	(B) Zero	(C) 1 to 3	(D) Four
		# years formally employed in the four years after graduation		
# students	11,799	19,405	22,822	20,873
Proportion of all students	0.16	0.26	0.30	0.28
Female	0.57	0.62	0.61	0.58
Age at graduation	23.90	23.71	24.16	24.20
College educated mother	0.38	0.28	0.30	0.28
Reputation	7.88	7.31	7.48	7.67
	(1.12)	(1.28)	(1.20)	(1.15)
Icfes	8.20	7.47	7.46	7.81
	(1.99)	(2.40)	(2.38)	(2.14)

Notes: The sample includes 2008–2009 graduates from the sample for Figure 1. We choose the 2008–2009 graduation cohorts so that we observe earnings for the first four years after graduation (2008–2011 for 2008 graduates, and 2009–2012 for 2009 graduates).

Column (A) includes any student who enrolled in a specialization, masters, or doctorate program in 2007–2011, the years for which we have graduate education records from the Ministry of Education. Column (B) contains non-graduate school students who never appear in our earnings records in the first four years after graduation. Column (C) contains non-graduate school students who appear in the earnings records in some but not all of the first four years. Column (D) contains non-graduate school students who appear in our earnings records in all four years.

Parentheses contain standard deviations. College educated mother is a dummy equal to one if a student's mother has a college/postgraduate degree.

Column (A) of Table B9 estimates (10) including Icfes terms but not reputation terms, such that the coefficients represent the unconditional returns to ability, a^u (equation (5), Section 2). The coefficient on Icfes shows that a ten percentile increase in the student's score is associated with a five percent increase in daily earnings in the year of graduation ($a_0^u \approx 0.05$). The standard deviation of Icfes percentiles is about twice that of reputation, and hence scaled by this measure the unconditional returns to reputation and ability are of a similar magnitude.

Proposition 2 states the coefficient on Icfes should increase with experience, i.e., it predicts a positive coefficient on the interaction of Icfes and experience. This follows from the assumption that employers do not fully observe Icfes scores, and thus the correlation of wages

all exam takers in the same test period based on a total Icfes score), but its percentiles take only integer values from one to 100.

Table B9. Returns to ability and experience interactions

	Dependent variable: log average daily earnings	
	(A)	(B)
Icfes	0.045	0.027
	(0.006)	(0.004)
Icfes $\times$ t	0.009	0.003
	(0.001)	(0.001)
N	83,492	83,492
R^2	0.163	0.297
# colleges	130	130
Extra controls		Y

Notes: The sample includes students in column (D) of Appendix Table B8 and earnings in the four years after graduation. Column (A) estimates equation (10) excluding reputation terms. In addition to the reported variables, the regression includes dummies for cohort-experience cells.

Column (B) adds the following controls to column (A): age at graduation, a gender dummy, dummies for eight mother's education categories, dummies for missing age and mother's education values, college program dummies, and dummies for college municipalities. Each control is interacted with a quadratic in experience.

Parentheses contain standard errors clustered at the college level.

and Icfes increases as workers reveal their skill through their output. Column (A) is consistent with this prediction. The point estimate on the Icfes-experience interaction implies that the return to ability grows by roughly 60 percent in the first four years after graduation.

Column (B) adds controls for graduates' gender, age, socioeconomic status, college program, and regional market. All controls are interacted with a quadratic in potential experience to allow earnings trajectories to vary with each characteristic. The coefficient on the Icfes-experience interaction decreases slightly, but it is still highly significant.

The increasing return to ability is similar to the Farber and Gibbons (1996) and Altonji and Pierret (2001) findings using AFQT scores as an unobserved characteristic. However, it is in contrast with findings in Arcidiacono, Bayer and Hizmo (2010), who also study AFQT scores but make a distinction between graduates who enter the labor market after high school and those who do so after college. For college graduates, they show that AFQT is strongly related to wages in the year of graduation, and this relationship changes little over the next ten years. Their conclusion is that AFQT revelation is complete for college graduates, and they suggest that this revelation occurs through college identity.

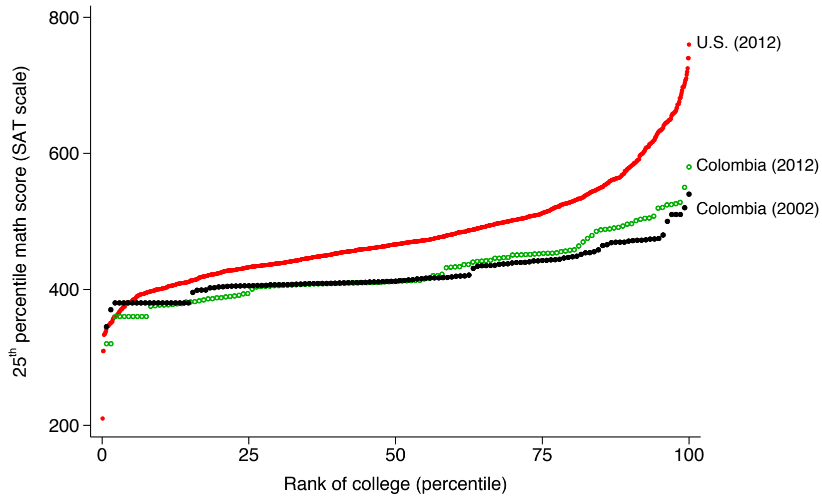


FIGURE B9. Ability sorting in Colombia and the U.S.

Notes: The y-axis shows the 25^{th} percentile math scores for entering students at U.S. and Colombian colleges. The x-axis depicts unweighted percentile ranks using these 25^{th} percentile math scores. U.S. SAT math percentiles are from the Integrated Postsecondary Education Data System. We include 1,271 four-year degree-granting public and private not-for-profit colleges with ten or more 2012 first-time degree/certificate-seeking undergraduates. Colombian colleges are the same as in Figure 1 (except three have no 2012 enrollees). We include students who enrolled in either 2002 or 2012 and took the Icfes no more than two years before enrolling. We calculate Icfes math percentiles relative to the enrollment cohorts and convert them to an SAT scale using the distribution of math scores for 2011 U.S. college-bound seniors, available in January 2015 at http://media.collegeboard.com/digitalServices/pdf/SAT-Mathemathics_Percentile_Ranks_2011.pdf. We jitter interior 25^{th} percentile math scores slightly to smooth out discrete jumps in SAT scores.

The difference in findings may be explained by the fact that sorting by ability in Colombia—although increasing—appears to be less extensive than in the U.S. Specifically, if the U.S. experience is indicative, one might expect sorting by ability to increase in Colombia as reductions in the cost of transport and information gradually move regional college markets away from relative autarky (Bound et al., 2009; Hoxby, 2009).

Figure B9 illustrates these dynamics in Colombia and its current standing relative to the U.S. We first plot the 25^{th} percentile Icfes math scores in the 2002 and 2012 entering cohorts at each college, with schools ranked on the x-axis according to this 25^{th} percentile.[126] To hold fixed the distribution of ability across cohorts, we use Icfes math percentiles relative to the population of college enrollees in the same year. For comparison with the U.S., we convert Icfes percentiles to an SAT scale using the distribution for 2011 college-bound seniors in the

[126] We plot the 25^{th} math percentiles for comparability with U.S. data. Other subjects and percentiles produce similar results.

U.S. There is evidence of increased sorting on math ability over the course of a decade. The top colleges in Colombia have experienced a 30 SAT point increase in their 25^{th} percentile scores, while the weakest have experienced a similar decline.

Despite these dynamics, by this measure Colombia's college market features substantially less sorting than that in the U.S. Figure B9 also shows the 25^{th} percentile SAT math scores for the 2012 entering cohort at U.S. four-year degree-granting public and private not-for-profit colleges. Comparing Icfes and SAT scores requires strong assumptions, as the tests may capture different characteristics, but 25^{th} percentile math scores increase much more rapidly in the U.S. While both countries have colleges with 25^{th} percentile scores below 400 SAT points, the top-ranked U.S. colleges are above 700, and no Colombian college surpasses 600.[127]

A plausible explanation for the positive coefficient on the interaction of Icfes and experience in Table B9 is thus incomplete sorting by ability across Colombian colleges. The more substantial sorting by ability across U.S. colleges may result in a more complete reflection of AFQT in wages upon graduation.[128]

B.10. **Return to years of schooling in Colombia.** Our main result from Section 4 is that the return to college reputation in Colombia increases with experience. This differs from the standard U.S. result that the return to years of schooling does not change with experience. This subsection shows that this benchmark years of schooling finding also holds in Colombia, as previewed in Panel A of Figure 4.

[127] If we convert Icfes scores to an SAT scale using the entire population of Icfes takers—instead of only those who entered college—the dots describing Colombia in Figure B9 shift up and become somewhat steeper, but they still exhibit a flatter slope than exists for U.S. colleges. This renormalization, however, overstates the amount of sorting in Colombia relative to the U.S. because Icfes test takers are less likely to enroll in college than SAT test takers. Using only college enrollees to make this conversion is more appropriate because the distribution of SAT scores we use is for U.S. college-bound seniors.

[128] If we estimate Table B9 with Icfes scores normalized to mean zero and standard deviation one—as Arcidiacono, Bayer and Hizmo (2010) do with AFQT—the period zero coefficient on Icfes is approximately one half of their AFQT coefficient. Although the two tests may measure different individual characteristics, the relative magnitudes are also consistent with partial revelation of the ability of college graduates in Colombia.

Table B10. Return to years of schooling and experience interaction

2008–2012 cross-sectional household survey

	(A)	(B)	(C)	(D)
	Dependent variable: Log hourly wage		Dependent variable: Log weekly earnings	
	0–39 years experience	0–9 years experience	0–39 years experience	0–9 years experience
Years of schooling	0.1224	0.1239	0.1150	0.1192
	(0.0008)	(0.0018)	(0.0009)	(0.0021)
Years of schooling × t	−0.0002	−0.0001	−0.0002	−0.0006
	(0.0000)	(0.0003)	(0.0000)	(0.0003)
N	660,573	217,523	660,573	217,523
R^2	0.407	0.352	0.351	0.308

Notes: Data for this table are from the 2008–2012 monthly waves of the Colombia Integrated Household Survey (*Gran Encuesta Integrada de Hogares*). The sample includes all workers who have hourly wages in the survey and 0–39 years of potential experience, t, which we define as $t = \min(\text{age} - \text{years of schooling} - 6, \text{age} - 17)$. Columns (B) and (D) restrict the sample to experience levels 0–9.

The dependent variable in columns (A)-(B) is log hourly wage. The dependent variable in columns (C)-(D) is log weekly earnings, defined as log hourly wage plus log usual hours of work per week.

In addition to the reported variables, all regressions include dummies for experience-year-month cells. Regressions are weighted by survey weights. Parentheses contain robust standard errors.

For this we use cross-sectional data from the 2008–2012 monthly waves of the Colombia Integrated Household Survey (*Gran Encuesta Integrada de Hogares*). This survey measures workers' hourly wages and years of schooling, which range from 0–20 years. We calculate each worker's potential experience, t, as $t = \min(\text{age} - \text{years of schooling} - 6, \text{age} - 17)$, and include workers with experience levels 0–39.[129]

Table B10 shows how the return to years of schooling in Colombia changes with experience. The use of cross-sectional data differentiates Table B10 from the panel data results in Farber and Gibbons (1996), Altonji and Pierret (2001), and Table 7 of this paper, but it is similar to the original Mincerian regressions that rely on U.S. survey data (e.g., Lemieux, 2006).

Column (A) displays the coefficients from a regression of log hourly wages on years of schooling and its interaction with experience.[130] The results suggest that an additional year

[129] We note that this definition of potential experience differs from the one we use elsewhere in the paper (earnings year minus graduation year) because the household survey does not include graduation dates. However, the age and schooling definition matches those in Altonji and Pierret (2001) and Lemieux (2006).

[130] Regressions in Table B10 also include controls for experience and survey date.

of education is associated with a 12 percent increase in initial wages, and that this gap remains roughly constant as workers gain experience. The coefficient on the interaction term is statistically significant due to the large sample size, but it is close to zero. For example, after ten years the return to schooling decreases by only 0.002 log points, or less than two percent of the initial return.

Column (B) of Table B10 restricts the sample to workers with 0–9 years of potential experience, with negligible impact on the results. This matches the experience levels we can observe using our administrative data on Colombian college graduates, as depicted in Panel B of Figure 4.

Columns (C)-(D) of Table B10 replicate columns (A)-(B) with log weekly earnings (rather than log hourly wage) as the dependent variable. This is motivated by the fact that we only observe earnings per day, not per hour, in our college administrative data. In both regressions, the coefficient on the interaction of schooling and experience remains close to zero. This suggests that the difference between the reputation and years of schooling findings is not driven our inability to observe hours worked.

In sum, the results of this subsection suggest that the standard Mincerian result of parallel earnings-experience profiles across schooling levels also holds in Colombia.

B.11. **Robustness of increasing return to reputation.** Table B11 documents the robustness of our main result from Section 4: the return to reputation—even conditional on Icfes scores—increases with experience (column (B) of Table 7). As a benchmark, we reproduce this result in column (A) of this table. The sample for this regression includes students from column (D) of Table B8. We regress log average daily earnings on dummies for cohort-experience cells, reputation, Icfes, and the interactions of both variables with experience. The point estimate on the reputation-experience interaction suggests that the effect of a one unit increase in reputation on earnings grows by about 1.2 percentage points each year.

Columns (B)-(D) test the sensitivity of this result to the addition of controls. Column (B) adds controls for gender, age at graduation, and socioeconomic status as measured

Table B11. Alternate specifications for return to reputation and experience interaction

	Dependent variable: log average daily earnings						
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
		Additional controls & experience interactions			Degrees of labor market attachment		
	Benchmark estimates	Gender, age, & SES	Program & municipality	Initial earnings	Actual experience	Full-time employment	No prior employment
Reputation	0.079	0.072	0.055	0.007	0.066	0.086	0.067
	(0.017)	(0.015)	(0.016)	(0.001)	(0.017)	(0.017)	(0.020)
Reputation × t	0.012	0.010	0.008	0.016	0.015	0.015	0.019
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)
Icfes	0.024	0.022	0.017	0.001	0.018	0.026	0.027
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.007)
Icfes × t	0.006	0.004	0.002	0.004	0.006	0.007	0.007
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
N	83,492	83,492	83,492	83,492	83,492	39,596	7,168
R^2	0.190	0.203	0.313	0.627	0.242	0.230	0.230
# colleges	130	130	130	130	130	130	113
Personal traits × $f(t)$		Y	Y	Y			
College traits × $f(t)$			Y	Y			
Initial earnings × $f(t)$				Y			
Definition of t	Potential	Potential	Potential	Potential	Actual	Potential	Potential
Full-time restriction						Y	Y
Prior work restriction							Y

Notes: All columns report coefficients on reputation, Icfes, and their interactions with experience. The sample includes the 2008–2009 graduates from column (D) of Appendix Table B8 and earnings within four years after graduation. All regressions include dummies for cohort-experience cells. Parentheses contain standard errors clustered at the college level.

Column (A) is identical to column (B) in Table 7. Columns (B)-(D) layer in additional controls, and every variable we add is interacted with a quadratic in experience. Column (B) adds a gender dummy, age at graduation, dummies for eight mother's education categories, and dummies for missing age and mother's education values. Column (C) includes all controls in column (B) plus program dummies and dummies for college municipalities. Column (D) includes all controls in column (C) plus log average daily earnings at experience zero.

Columns (E) is identical to column (A), but all experience terms are defined using actual experience—the cumulative number of months with earnings since graduation—rather than potential experience. Column (F) is identical to column (A), but we include only graduates who have earnings in every month starting in the year after graduation. Column (G) includes only those students in column (F) who do not appear in our earnings records in the year prior to graduation. This column includes only 2009 graduates, for whom we can observe pre-graduation employment.

by mother's education. We interact all variables with a quadratic in experience so that controls can affect both the intercept and the slope of graduates' earnings profiles. The addition of these controls for personal characteristics lowers the coefficient on the interaction of reputation and experience slightly, though it is still significant and roughly the same magnitude in proportion to the period-zero return to reputation.

Column (C) includes all controls from column (B) and adds two characteristics of graduates' colleges. First, we add dummies for college programs (see column (C) of Table B1) and their interaction with a quadratic in experience. These dummies are important if graduates from different programs enter occupations that vary in their potential for wage growth. Second, we add dummies for college municipalities and the interactions of these dummies with an experience quadratic. Location controls may matter if earnings paths differ across regional markets. Our estimates in column (C) are thus identified off of variation in college reputation for students in the same programs and cities. The magnitude of the reputation-experience coefficients falls again, but it is still significant and is slightly larger in relation to the initial return to reputation.

In addition to the controls in column (C), column (D) adds each graduate's log earnings in the year of graduation. The inclusion of experience-zero earnings is in the spirit of Farber and Gibbons (1996), who use initial wages to control for other worker characteristics observable to employers but not to the econometrician. We additionally interact initial earnings with a quadratic in experience to control for variation in earnings trajectories across jobs with different starting wages. The controls for initial earnings mechanically reduce the period-zero reputation and Icfes coefficients, but the coefficient on the interaction of reputation and experience doubles in magnitude relative to column (C).

In columns (E)-(G), we remove the controls from columns (B)-(D) and instead test the sensitivity of our result to the degree of graduates' labor market attachment. As discussed, the sample for Table B11 includes only students who are employed in each of the first four years after graduation, but graduates may still differ in the number of months they are employed in each year. In all previous specifications, we measure labor market experience using potential experience, defined as calendar year minus graduation year. Column (E) of Table B11 is identical to column (A), but we replace all experience terms with *actual* experience, defined as the number of months of employment since graduation.[131] This alternate measure may be important if graduates from high reputation colleges are more likely to find stable employment, but the results in column (E) are similar to our benchmark estimates.

Column (F) is identical to column (A), but we restrict the sample to include only students who have full-time employment after graduation. In column (A) we require that each student have at least one monthly earnings observation in each of the first four years after graduation. In column (F), students must have an earnings observation in *every* month beginning in the year after graduation. This requirement reduces the sample size by more than 50 percent but has little effect on the reputation-experience coefficient.

Column (G) makes a further restriction to the sample from column (F). In this column we also require that graduates were not employed in the year *before* graduation. This restriction may be important if graduates from top colleges are less likely to work while in school, and if prior employment affects future wage growth. Since our earnings records begin in 2008, we can only observe pre-graduation employment for 2009 graduates. Thus, column (G) includes only 2009 graduates who have no earnings in 2008. This restriction leads to a small sample in column (G), but if anything, the coefficient on the interaction of reputation and experience is larger in this population.

In sum, Table B11 suggests that the increasing conditional return to reputation is not driven by variation in earnings paths across individual characteristics, college programs, regional markets, or levels of initial earnings. Furthermore, this result does not appear to stem from variation across colleges in labor market attachment.

[131] Papers in the employer learning literature use different measures of experience and potential experience. Farber and Gibbons (1996) use experience based on actual employment duration, while Altonji and Pierret (2001) principally use potential experience based on age and years of schooling. Potential experience based on graduation year is most logical for our study of college reputation and is consistent with the primary measure used by Arcidiacono, Bayer and Hizmo (2010).

TABLE B12. Experience interactions with college-program level reputation

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)
Reputation	0.064	0.039	0.044	0.024
	(0.015)	(0.014)	(0.014)	(0.013)
Reputation $\times$ t	0.012	0.012	0.008	0.004
	(0.002)	(0.002)	(0.002)	(0.001)
Icfes		0.045	0.034	0.023
		(0.005)	(0.004)	(0.004)
Icfes $\times$ t			0.007	0.002
			(0.001)	(0.001)
N	83,492	83,492	83,492	83,492
R^2	0.156	0.178	0.179	0.301
# colleges	130	130	130	130
Extra controls				Y

Notes: This table is identical to Table 7 in Section 4, but it uses reputation defined as mean Icfes at the college-program level rather than at the college level. The dependent variable is log average daily earnings. The sample includes students in column (D) of Appendix Table B8 and earnings in the four years after graduation. Columns (A)-(C) estimate equation (10) excluding and including Icfes terms. In addition to the reported variables, both regressions include dummies for cohort-experience cells.

Column (D) adds the following controls to column (C): age at graduation, a gender dummy, dummies for eight mother's education categories, dummies for missing age and mother's education values, college program dummies, and dummies for college municipalities. Each control is interacted with a quadratic in experience.

Parentheses contain standard errors clustered at the college level.

B.12. **College-program level reputation.** This section provides details on the robustness of our main results using a college-program level definition of reputation rather than a college level definition. Table B13 replicates the main exit exam results from Table 3 with reputation defined as college-program mean Icfes. The results closely mirror our main findings in sign and magnitude, although the standard errors are typically larger. This is likely due to the fact that the college-program reputations are calculated from smaller samples. In general this does not alter the pattern of statistical significance relative to Table 3, with the exception of statistically insignificant reputation effects in columns (B) and (D).

Table B12 replicates the results on earnings growth from Table 7 in Section 4 with college-program level reputation. The main findings are unaltered by this modification. In particular, the coefficient on the reputation-experience interaction is positive and highly significant in all specifications.

Table B13. Exit exam effects with college-program level reputation

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)	(F)
		Experience & cohort controls		Restriction to similar programs		
	Benchmark specification	Within experience	Linear trends	S. sciences & engineering	Within $\hat{r}_p$ quartiles	Within $\hat{a}_p$ quartiles
Reputation × δ_{pc}	−0.038	−0.021	−0.015	−0.052	−0.026	−0.054
	(0.022)	(0.019)	(0.028)	(0.031)	(0.012)	(0.023)
Icfes × δ_{pc}	0.019	0.017	0.010	0.043	0.014	0.009
	(0.007)	(0.009)	(0.009)	(0.012)	(0.007)	(0.004)
N	581,802	267,924	267,924	273,590	581,802	581,802
R^2	0.258	0.225	0.225	0.263	0.259	0.259
# programs	39	39	39	22	39	39
Experience levels	0–9	4–7	4–7	0–9	0–9	0–9

Notes: This table is identical to Table 3 in Section 3, but it uses reputation defined as mean Icfes at the college-program level rather than at the college level. All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. Regressions in columns (A) and (C)-(F) include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. Column (B) includes dummies for program-cohort-experience cells and interactions of both reputation and Icfes with program-experience and cohort-experience dummies. The sample for each regression is restricted to the experience levels listed in the bottom row. Parentheses contain standard errors clustered at the program level.

Column (C) adds interactions of both linear experience and cohort terms with college reputation and Icfes for each program. Column (D) restricts the sample to social sciences and engineering program areas and adds interactions of dummies for social-science-area-cohort cells with both reputation and Icfes. Column (E) adds interactions of both reputation and Icfes with dummies for cells defined by cohort and each program's quartile of the returns to reputation estimated from 2003–2004 cohorts. Column (F) adds interactions of both reputation and Icfes with dummies for cells defined by cohort and each program's quartile of the returns to Icfes estimated from 2003–2004 cohorts.

Chapter 3. Time Gaps in Academic Careers

1. Introduction

A defining feature of the U.S. education system is that it permits second chances. Many European countries have national exams and centralized admissions that assign students to educational tracks, often at young ages. In the U.S., graduation standards have historically been left up to local districts, and colleges vary widely in admission criteria. U.S. students can thus more readily switch academic paths, or return to school after dropping out.

A second-chance system helped the U.S. become an early leader in educational attainment, but it may also partly explain the recent stagnation of high school and college graduation rates (Goldin and Katz, 2008*b*). Lenient standards can weaken accountability and reduce school productivity. This concern has motivated research on accountability systems (Kane and Staiger, 2002; Hanushek and Raymond, 2005; Figlio and Loeb, 2011), teacher evaluation (Rivkin, Hanushek and Kain, 2005; Chetty, Friedman and Rockoff, 2014), and school choice (Hoxby, 2003; Urquiola, 2016).

This paper explores a different channel through which flexible standards affect academic outcomes. Forgiving education systems create churning by allowing students to defer the completion of their schooling. The result is that many students experience time gaps in their academic careers as they cycle in and out of school. In the U.S., for example, roughly one-third of all students who matriculate in college wait more than one semester after high school to enroll. Such gaps, as measured by age differences between secondary graduates and tertiary enrollees, tend to be larger in OECD countries without educational tracking.

I examine whether academic time gaps can lower schooling attainment. This builds on research that asks how the structure of education systems—e.g., school entry or exit ages—affects attainment (Angrist and Krueger, 1992; Bedard and Dhuey, 2006; Dobkin and Ferreira, 2010; Black, Devereux and Salvanes, 2011; Fredriksson and Öckert, 2014). My focus is on the potential gaps that arise in the transition from high school to college. Individuals' experiences during these gaps may affect their decision of whether to acquire further education at all. Joining the labor force can reveal the job satisfaction and earning potential of an existing degree. Time away from school may make it harder to go back.

Isolating the effect of academic gaps on subsequent enrollment requires quasi-random variation in transition timing in an education system. For this I use administrative data from the country of Colombia and a policy that altered its unique system of academic calendars.

Colombian high schools operate on two distinct annual schedules; some schools begin the academic year in January, while others start in September. All public high schools operate on the January calendar except for those in two regions, which historically began in September. From 2008–2010, these regions transitioned public schools to the January calendar. This altered the academic term at nearly 400 high schools that I call *switching schools*, which include all public schools in the affected regions plus some local private schools that followed suit. Other private high schools in these regions did not switch calendars. I use the term *staying schools* to refer to 89 local private schools that stayed on the September calendar.

Colombian colleges offer admission every semester, so students on either high school calendar can typically begin college right after graduation. But the calendar transition led to an unusual gap before potential college entry at both switching and staying schools, for separate reasons. Switching schools transitioned to the new calendar by adding mid-term breaks and delaying graduation by a few months each year. This caused one cohort—2009 graduates—to finish high school just after the start of the September semester at most colleges. Thus, many 2009 graduates could not enter college until January—one semester later than was typical for prior cohorts at switching schools.

Students from staying schools were also affected by the transition. Colombian students apply to both a college and a major, and while many programs are offered twice per year, some are annual. To align with the new public high school calendar, colleges in the affected regions shifted some annual majors from a September to a January start date. Graduates from staying schools who were interested in these programs therefore had to wait an extra semester to enroll. This yielded a second source of post-graduation time gaps, induced in this case by changes in college calendars rather than high school calendars.

At both switching and staying schools, the calendar changes led to sharp declines in the number of students who began college in the first semester after graduation. I show this in a differences in differences analysis with students in unaffected regions as a comparison group. This initial enrollment reduction is a first stage result; the calendar shift provides quasi-random variation in the occurrence of one semester gap between high school and college.

I then consider how this time gap affects subsequent college enrollment. Students who did not enroll in college within one semester could typically have done so within one year. Many did not. Roughly 50 percent of those exposed to the time gap did not enroll in the next possible semester. This occurred at both switching and staying schools, leading to a seven percent reduction in college attendance rates in affected regions. Further, there is little catch-up enrollment in the next two years, suggesting that the decline is not merely temporary.

The decline in college enrollment is evident graphically and in regressions, and it is robust to different comparison groups and to the inclusion of linear trends for each high school. Staying schools provide a clean test of the timing effect because their graduates had few changes in college preparation. Graduates from switching schools were affected by other elements of the transition, such as a longer academic year. But I find little evidence that these other channels are important; the results do not appear to be driven by changes in graduates' entrance exam scores or admission rates at selective colleges.

I explore two sources of heterogeneity that inform the main results. First, I find that low SES and low ability students are more likely to forgo college after the time gap. These students also had lower college enrollment rates in the pre-policy years. Thus time gaps may have a greater effect on individuals who are indecisive about college in the first place.

Second, I show that the enrollment decline comes from students forgoing low-paying majors; those who entered college after the gap tended to choose high-wage majors. This could arise if working or job-hunting in the interim altered students' relative valuations of different degrees. Consistent with this, I use household survey data to show that labor force participation increased among 17 year olds in the affected regions during the gap period.

The bottom line is that time gaps in education systems can affect both the mean and the distribution of schooling, with potential implications for a nation's wage growth and inequality (Goldin and Katz, 2008*b*; Acemoglu and Autor, 2011).

My paper relates to four research areas in labor and education economics. First, it contributes to work on institutional features that affect transitions between education tiers. Some researchers have argued that the U.S.'s disconnected education system, in which K–12 and postsecondary schools function as separate entities, can hinder student outcomes (e.g., Venezia, Kirst and Antonio, 2003). This concern underlies research on the benefits of educational tracking (Hanushek and Woessmann, 2006; Schütz, Ursprung and Woessmann, 2008) and remediation classes (Bettinger and Long, 2009; Martorell and McFarlin Jr, 2011).

I show that timing is another potential disconnect in the transition from high school to college. This adds to work on the "summer melt," a phenomenon in which students who are planning to attend college at the time of high school graduation change their minds by the end of the summer (Castleman and Page, 2014). My findings suggest that tracking or curriculum policies can have additional attainment benefits if they limit time gaps in students' academic careers. Further, policies that promote college access may have a greater impact if they are implemented before students finish compulsory education.

Second, this paper adds to research on the information hurdles that affect schooling choices. Recent work finds that students' decisions respond to information on tuition costs and financial aid (Bettinger et al., 2012; Hoxby and Turner, 2013*b*; Dynarski and Scott-Clayton, 2013), returns to education (Jensen, 2010; Nguyen, 2010; Oreopoulos and Dunn, 2013; Dinkelman and Martínez A., 2014), and school test scores (Hastings and Weinstein, 2008).

In my paper, the information that alters enrollment decisions is acquired not through interventions but by experiences outside of school. This is consistent with Perez-Arce (2015), who finds that applicants with deferred admission to a Mexican university are more likely to forgo college if they work in the interim. Goodman (2013) also shows that the experience of taking the American College Test (ACT) induces some students to attend college. More broadly, my findings contribute to work on how routines/defaults (Scott-Clayton, 2011; Pallais, 2015) or present biases (Angrist and Lavy, 2009; Angrist, Lang and Oreopoulos, 2009; Fryer, 2011) affect educational choices, which may explain the time gap effects.

Third, recent research finds that students' choice of major depends on expected earnings (Arcidiacono, Hotz and Kang, 2012; Hastings, Neilson and Zimmerman, 2015), initial beliefs (Stinebrickner and Stinebrickner, 2014), and heterogenous tastes (Arcidiacono, 2004; Beffy, Fougere and Maurel, 2012; Wiswall and Zafar, 2015). I present evidence that program choices change when individuals spend time outside of the education system. This finding parallels that in Zafar (2011), who shows that students update their major preferences based on their performance in college.

Finally, my paper contributes to work on education externalities from labor market conditions. Booms arising from coal prices (Black, McKinnish and Sanders, 2005), trade reforms (Atkin, 2012), and the housing market (Charles, Hurst and Notowidigdo, 2015) can lead individuals to forgo further schooling. My results suggest that education externalities can also arise from an individual's labor force participation prior to the completion of her academic career.

The next section gives background on the prevalence of time gaps between high school and college in different education systems. Section 3 describes the academic calendar shift in Colombia. Section 4 discusses the resulting effects on schools that switched to the January schedule. Section 5 shows the effects on schools that stayed on the September calendar. Section 6 explores heterogeneity in these results at both school types. Section 7 concludes.

2. Time gaps between high school and college

Every spring roughly three million U.S. students receive their high school diplomas. Approximately two-thirds of them enroll in a two- or four-year college in the fall of the same year. This ratio is sometimes called the *immediate college enrollment rate*, and in the U.S. it has hovered between 60 and 70 percent for the past two decades.[132]

For graduates who do not immediately attend college, the decision to forgo further education is not necessarily a permanent one. The dashed line in Figure 1 illustrates this using the 1997 National Longitudinal Survey of Youth (NLSY). It plots the cumulative fraction of graduates who have ever enrolled in college against the number of years since their high school graduation. The immediate college enrollment rate for NLSY graduates is slightly below 60 percent, as indicated by the data point at 0.5 years after graduation. The fraction ever enrolled rises further after this initial spike, reaching nearly 80 percent after nine years. This value reflects the *overall college enrollment rate*, which would be higher if the data included enrollment at ten or more years.[133]

The solid line in Figure 1 replicates this curve for Colombian high school graduates using administrative data described below. Both immediate and overall college enrollment are substantially lower in Colombia than in the U.S.

Figure 1 shows that students in both countries experience time gaps between high school and college. In the U.S., 30 percent of students who eventually attended college did not enroll in the first semester after graduation. Delayed enrollment is even more common in Colombia, comprising more than two-thirds of all college-goers. Further, Appendix A.1 shows that these time gaps vary with students' backgrounds. Low SES and low ability students are more likely to delay college entry, as are students who are older than the normative graduation age. This pattern holds in both the U.S. and Colombia.[134]

Figure 2 explores how post-secondary gaps vary across education systems. Since longitudinal data like that in Figure 1 do not exist in some countries, it uses age gaps as a proxy for time gaps. Specifically, it plots the difference between the average ages of new tertiary

[132] This statistic is from the National Center for Education Statistics (NCES) (available in December 2015 at http://nces.ed.gov/programs/coe/indicator_cpa.asp).

[133] The overall enrollment could be substantially higher; an NCES report finds that 12 percent of the 1995 incoming cohort waited ten or more years to enroll (available in December 2015 at http://nces.ed.gov/pubs2005/2005152.pdf, Figure I).

[134] I find that gender is not a significant predictor of delayed enrollment in either country.

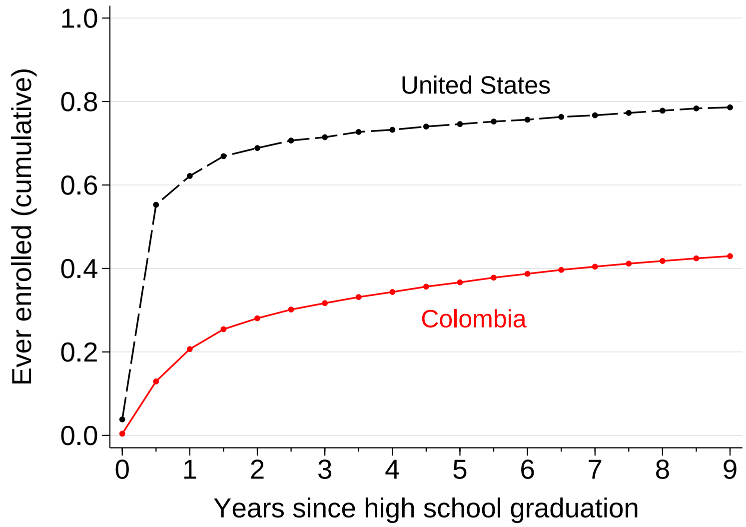


Figure 1. College enrollment timing

Notes: The U.S. sample is students in the 1997 NLSY cross-sectional and over-samples with a high school graduation year in 1995–2002. The x-axis is the difference between college enrollment and high school graduation dates, aggregated into six month bins. Averages use 2011 panel weights.

The Colombian sample is any 11^{th} grader who took the national college entrance exam in 2001–2002. The x-axis is the difference between the semesters of college enrollment and the exam. See Section 3 for details on the Colombian data, entrance exam, and academic calendar structure.

For both countries, I count a student as ever enrolled if she enters a two- or four-year college in the dataset within nine years of her graduation month. Enrollment at zero years includes students reported to begin college in or before the month of high school graduation.

entrants and secondary graduates for OECD countries with data in 2011 (plus Colombia).[135] Following Hanushek and Woessmann (2006), I classify countries based on early age tracking of students into different school types. Light grey bars depict countries with tracking below age 16; black bars indicate countries without tracking by this age.

Age gaps between secondary and tertiary education vary significantly across countries, ranging from below one to nearly six years. Further, countries with tracking systems tend to have smaller age gaps. The smallest gaps occur in countries with early tracking; these include Belgium, the Czech Republic, Germany, and Mexico, all of which begin tracking by age 12. Colombia, the U.S., and Scandinavian countries do not have tracking systems and are near or above the median. Other institutional features also affect age gaps. The large gap

[135] Figure 2 includes only students enrolled in the highest tiers of secondary (general programs) and tertiary (level A, further education/theoretically based) school. It does not therefore capture age gaps for students pursuing technical education, which may vary across countries. The sample also includes only students below age 30; without this restriction the gaps would be larger, and the country ranking may change. Note that the Colombian gap is calculated from the data described in Section 3, which may affect comparability.

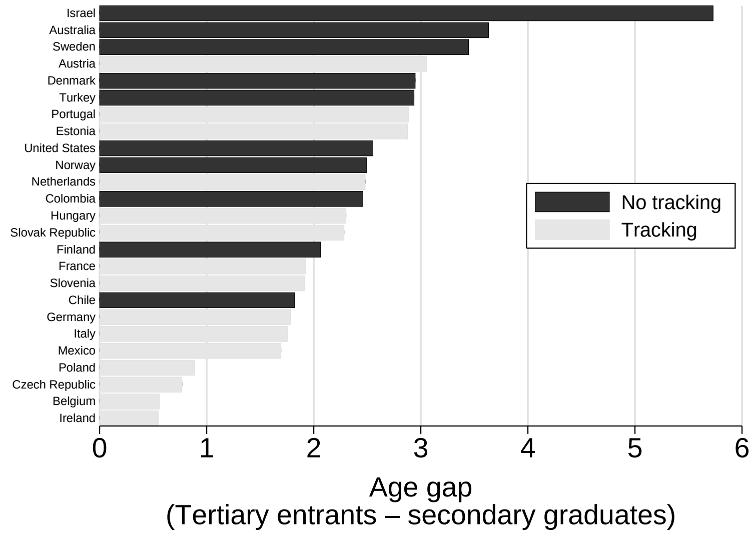


FIGURE 2. OECD countries — Secondary/tertiary age gap

Notes: Data are from the OECD for the year 2011 (available in December 2015 at http://stats.oecd.org). Bars depict the difference between the average ages of new entrants to level A tertiary programs and graduates from upper secondary general programs. In the U.S., upper secondary includes all education programs because there is no distinction between general and vocational/technical programs. In Colombia, the age difference is between entrants to university-level programs and 11th grade college entrance exam takers in 2011 using the data described in Section 3. All secondary and tertiary averages are calculated using students age 29 or below, for which single-year ages are available from the OECD.

Light grey bars depict countries that track students into school types before age 16. This classification follows Hanushek and Woessmann (2006) and their cited source (available in December 2015 at http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/34002216.pdf, Figure 5.20a). Countries not in this source are classified as in Education Policy Outlook 2015 (available in December 2015 at http://www.oecd.org/publications/education-policy-outlook-2015-9789264225442-en.htm).

in Israel is due in part to its military service requirement. In Ireland—the country with the smallest gap—most secondary students participate in a "transition year" of non-academic subjects and volunteer work before graduating.

Figure 2 suggests that different education systems may lead to substantial variation in the existence of gaps between schooling tiers. These gaps can affect overall tertiary enrollment rates if experiences outside of the education system alter students' views about the returns or costs to further schooling. In the U.S., roughly three-quarters of high school graduates who do not immediately enter college join the labor force by the fall of the same year.[136] Employment or job-hunting may alter the perceived value of students' existing education.

[136] This statistic is for 2014 graduates and comes from the Bureau of Labor Statistics (available in December 2015 at http://www.bls.gov/news.release/hsgec.nr0.htm).

The gaps in Figure 2 are likely correlated with other cultural or labor market factors that affect college attendance rates. Isolating the time gap effect on subsequent enrollment requires exogenous variation in delays induced by an education system. In the next section, I describe a policy that altered Colombia's academic calendars that provides an opportunity to test for the causal effect of time gaps after high school.

3. An academic calendar shift in Colombia

This section describes the high school and college calendars in Colombia, my related data sources, and a government policy that altered these calendars in two regions of the country.

3.1. **Academic calendars and college admissions.** The high school system in Colombia is unique in that students begin the school year at two different times. The large majority of schools start the academic year in January and conclude in November. This schedule yields the longest break during the Christmas season. A small number of private high schools begin in September and finish in June. These schools choose a calendar that aligns with that of U.S. and European colleges.

To enter college, Colombian students are required to take a national standardized admission exam called the Icfes.[137] The exam is generally analogous to the SAT in the U.S., but its results are also used to evaluate high schools. As a result, the vast majority of graduates take the exam, even those who do not plan to attend college.[138] The Icfes is offered semiannually at the end of the last year of high school on each calendar. Scores are returned promptly, and students who apply to college can start in the next semester.

Accordingly, Colombian colleges also enroll students two times per year. Students can begin in either January or September, and nearly all colleges offer admission in both semesters.[139] Unlike the U.S., applicants choose both schools and majors at the time of application, but

[137] Icfes stands for Institute for the Promotion of Higher Education, the former acronym for the agency that administers the exam. The Icfes exam is now named Saber 11°, but I use the name Icfes to match the designation during most of the period covered by my data.
[138] The exam agency estimates that over 90 percent of graduates take the exam.
[139] The January cohort is slightly larger at many colleges because most high schools use the January calendar, but the cohort sizes are more balanced than in the U.S., where almost all undergraduates begin in the fall.

the Colombian market is similar in that there are many public and private colleges with varying selectivity and degree durations. Admissions are also decentralized; students must apply to individual colleges, and each institution controls its own selection criteria.

3.2. **Data sources.** This paper uses two administrative data sources:

(1) Records for all high school seniors (11^{th} graders) who took the Icfes college entrance exam from 2001–2011. These data are from the testing agency and contain each student's high school, background characteristics, and exam scores.
(2) Records for students enrolling in college in 2001–2012. These are from the Ministry of Education and cover almost all higher education institutions in Colombia.[140] The records include each student's enrollment date, program of study, and institution.

I merge these two datasets using national ID numbers, birth dates, and names. This defines the measure of college enrollment for my analysis, which is an indicator for appearing in the Ministry of Education records.[141]

In addition, I use two datasets related to students' labor market outcomes. First, administrative data from the Ministry of Social Protection provides 2008–2012 earnings for all college enrollees employed in the formal sector.[142] I use these data to calculate the average earnings of college graduates by major. Second, I use 2007–2010 Colombian household survey data to measure labor force participation by age and region.[143]

3.3. **An academic calendar shift.** Almost all public high schools in Colombia begin the academic year in January, but in two regions the public school year historically started in September. These regions—Valle del Cauca and Nariño—are the third and eighth largest of Colombia's 33 administrative departments, and the capital of Valle del Cauca, Cali, is the country's third most populous city. I call these two departments the *affected regions.*

[140] See Appendix A.3 for details on data coverage.
[141] I match over 91 percent of college enrollees who took the Icfes during the period covered by the data. Appendix A.2 provides details on the merge and match rates.
[142] My administrative earnings records include only college enrollees, so I cannot use these data to examine labor market outcomes of individuals who do not attend college.
[143] The survey is the *Gran Encuesta Integrada de Hogares* (GEIH).

Anecdotally, the affected regions contained some of Colombia's first private high schools, and thus public schools adopted the same September calendar.

In 2008, the affected regions transitioned public schools to the January calendar to align with the federal fiscal year and with all other public schools. The transition occurred over two years by adding extra instructional periods and mid-term holidays (see Figure 3 below for details). The academic calendar shift was complete by January 2011.

Although the policy affected public high schools, some local private schools also changed from the September to the January calendar. Other private schools chose not to change schedules. I refer to the public and private schools that switched calendars as *switching schools*, and the private schools that stayed on the September calendar as *staying schools*.

Columns (A)–(C) in Table 1 describe switching and staying schools in the affected regions. Switching schools include all 290 public schools plus 109 private schools, while 89 private schools stayed on the September calendar.[144] Staying schools are higher quality by several metrics. 97 percent of staying schools received the exam agency's high or superior rank in any year, compared with half of switching schools. 55 percent of staying schools offer academic-level training, while two-thirds of switching schools provide technical education.

Table 1 also shows the characteristics of 2001–2011 graduates from switching and staying schools. There are roughly 25,000 switching school graduates per year, and their average Icfes exam score is near the median of the national distribution. Staying schools graduate less than 4,000 students per cohort, and their average student performs at the 74$^{\text{th}}$ percentile on the Icfes. Staying school graduates also have higher socioeconomic backgrounds as measured by mother's education.

Lastly, Table 1 shows the fraction of graduates who enroll in college, defined as appearing in any institution in the Ministry of Education records. Roughly 10 percent of switching school graduates enroll in the semester immediately after the Icfes exam. This fraction

[144] Table 1 does not contain all high schools in Colombia because I include only those with exam takers in all years of my analysis (2001–2011). I also omit schools that ever had a "flexible" calendar, in which students can begin the school year in either semester. See Appendix A.4 for details on the high school sample.

Table 1. Summary statistics by school type

	(A)	(B)	(C)	(D)	(E)
	Affected regions			Other regions	
	Switching schools		Staying schools		
	Public	Private	Private	Public	Private
High school characteristics					
# high schools	290	109	89	2,553	1,156
High/superior rank	0.47	0.58	0.97	0.57	0.90
Academic school	0.38	0.33	0.55	0.50	0.66
2001–2011 graduate characteristics					
Total students per year	19,862	5,247	3,727	174,335	56,360
Icfes percentile	0.51	0.54	0.74	0.49	0.67
Mother attended college	0.12	0.26	0.65	0.14	0.49
Younger than 18	0.66	0.76	0.72	0.72	0.82
College enrollment by years since Icfes					
Enrolled within 0.5 years	0.09	0.13	0.28	0.09	0.31
Enrolled within 3 years	0.26	0.35	0.55	0.29	0.60
Enrolled within 6 years	0.33	0.41	0.59	0.37	0.66

Notes: The sample is 11^{th} graders who took the Icfes in 2001–2011 and attended high schools with exam takers in every year. Columns (A)–(C) show high schools in Nariño and Valle del Cauca; columns (D)–(E) include high schools in all other regions. Switching schools are those on the January calendar in 2010 and/or 2011, and on the September calendar in all other years. Staying schools are on the September calendar in all years. I omit high schools that are ever listed with a "flexible" academic calendar, affected region schools that change calendars before 2010, and schools in other regions that ever change calendars. Private schools are those that are listed as private in any year; public schools are listed as public in all years.

A school is high/superior rank if it ever received one of the exam agency's top three (of seven) ranks in 2001–2008. A school is academic if it is academic or normal in any of these years, and zero if it is always technical or academic & technical. Averages of these variables are weighted by the number of exam takers.

Icfes percentiles are relative to all 11^{th} grade exam takers in the same year and are calculated using the average of the scores from the six core components that did not change in 2001–2011: biology, chemistry, language, mathematics, philosophy, and physics. Mother attended college equals one if a graduate's mother has any degree above basic secondary; this variable is only available for the 2008–2011 cohorts. Age is calculated at the end of August in the exam year.

College enrollment within t years of the Icfes is defined as a graduate's first appearance in the Ministry of Education college records. These averages are calculated using 2001–2006 graduates.

See Appendices A.3 and A.4 for further details on the included colleges and high schools.

rises to roughly 40 percent by six years after the exam. Both immediate and overall college enrollment are substantially more common in staying schools.

The next two sections explore the effects of the calendar shift on college enrollment using schools in other regions of Colombia as a comparison group (columns (D) and (E) in Table 1). In particular, the calendar transition created a time gap between high school and college

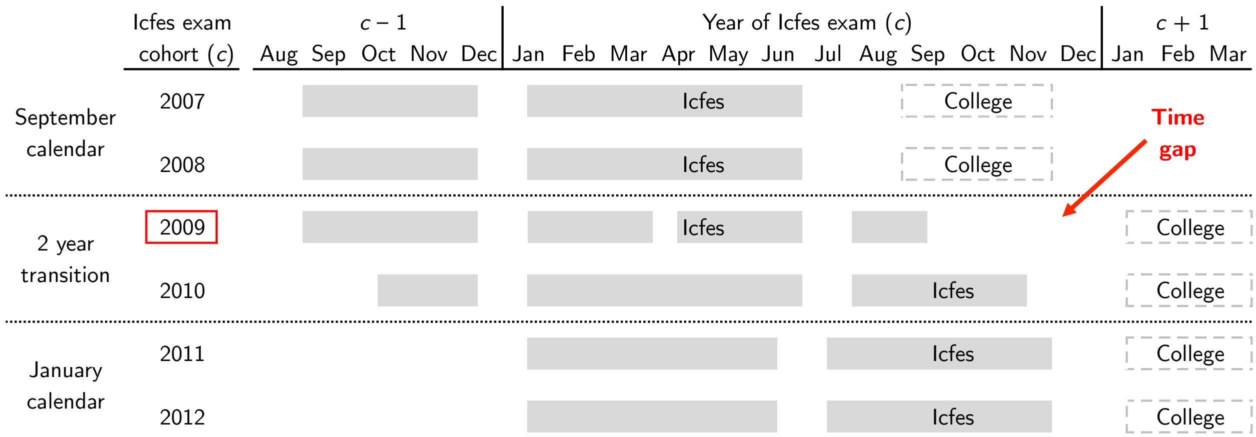


FIGURE 3. Calendar transition and a time gap

Notes: Grey bars are instructional periods in the last year of high school. Gaps are break periods. White boxes represent the first potential college semester. Icfes indicates the timing of the college entrance exam.

Schedules are approximate based on half-month increments; there are small yearly differences in start/end dates and Icfes timing. This figure is based on 2006–2012 resolutions from the Secretary of Education in Cali, Valle del Cauca. Resolutions from the Security of Education in Nariño show similar schedules.

at both switching schools and staying schools, but for different reasons. The next section focuses on switching schools; Section 5 discusses staying schools.

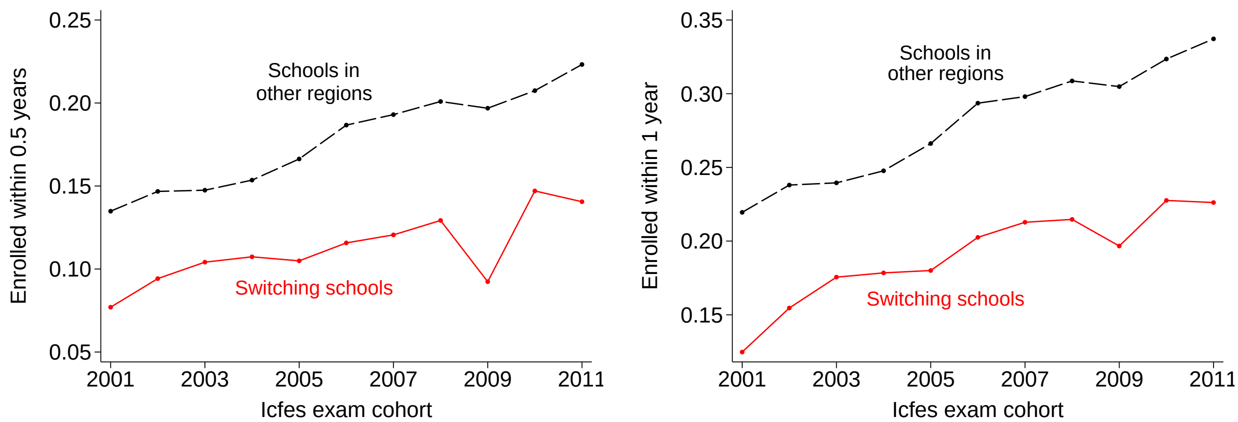
4. Switching schools

This section describes how the calendar transition led to a time gap after graduation for one cohort at switching schools. It then presents the resulting effects on college enrollment.

4.1. **Calendar transition and a time gap.** Figure 3 illustrates the transition from the September to the January academic calendar at switching schools.[145] It depicts the 2007–2012 cohorts, defined by the year of the Icfes college entrance exam. The grey bars represent instructional periods in the last year of high school. Prior to 2009, students began the year in September, took the Icfes exam in April/May, and graduated in June. They were eligible to begin college in September of their exam year, indicated by the white boxes.

2009 graduates started senior year and took the Icfes on the typical schedule, but they graduated three months later than usual. This occurred through one extra month of classes and two additional mid-term breaks. 2010 graduates also had an extra month of instruction,

[145] Figure 3 depicts academic calendars from public resolutions by the government in Cali, Valle del Cauca. Resolutions from Nariño show a similar proposed transition.



Panel A. Enrolled within 0.5 years

Panel B. Enrolled within 1 year

FIGURE 4. Switching schools — College enrollment by years since Icfes

Notes: Switching schools include students from columns (A) and (B) in Table 1. Schools in other regions include students from columns (D) and (E). College enrollment is defined as in Table 1.

but their school year began and ended several months later. The transition was complete by 2011, when the academic calendar matched that for all other Colombian public schools.

Figure 3 shows that the transition shifted the first possible college enrollment semester from September to January. This created an unusual four month gap between high school and potential college enrollment for the 2009 cohort, which graduated just after the September college semester began. Below I ask how this time gap altered the decision to enroll in college. I also explore other potential confounding effects such as changes in class time.

4.2. **Benchmark effects on college enrollment.** To study the effects of the time gap, I compare college enrollment among graduates from switching schools (columns (A) and (B) in Table 1) to graduates from other regions of Colombia (columns (D) and (E)).

Figure 4 plots the mean college enrollment rates in switching schools and other schools by Icfes exam cohort. Panel A shows the fraction of graduates entering college one semester after taking the Icfes. There is a sharp drop in the 2009 switching school cohort. This decline in enrollment is in the September college semester, which began before many of these students graduated (see Figure 3). The 2009 enrollment rate is not zero, suggesting that some students or schools graduated early. Nonetheless, many 2009 graduates did not enter college as usual.

The initial decrease in college enrollment in Panel A is a "first stage" result; some 2009 graduates had to wait one extra semester to enroll. Panel B explores the effects of this time gap on enrollment within one year of the exam—the next possible college entrance date. The enrollment decline for the 2009 cohort is less pronounced but still evident; the drop is roughly half of its magnitude from Panel A. This suggests that many students did not enter college following the one semester gap.

Table 2 quantifies these effects using the differences in differences regression

$$y_{hc}^t = \gamma_h + \gamma_c + \beta^t \delta_{hc} + \epsilon_{hc}. \tag{1}$$

The dependent variable, y_{hc}^t, is the fraction of students from high school h and Icfes exam cohort c who enroll in college t years after the exam. The regression includes high school dummies, γ_h, cohort dummies, γ_c, and a treatment indicator, δ_{hc}, which equals one for the 2009 cohort at switching schools. The coefficient of interest, β^t, measures the change in enrollment within t years for 2009 switching school graduates relative to other schools. This is a school-cohort level regression with observations weighted by the number of exam takers.[146]

Panel A of Table 2 displays the β^t coefficients from (1). The columns use dependent variables that measure cumulative enrollment at different durations t. The sample includes only 2001–2009 graduates, for whom I observe enrollment up to $t = 3$ years later.

Column (A) shows that for 2009 switching school graduates, college enrollment in the semester after Icfes declined by about five percentage points. This mirrors the "first stage" result from Panel A of Figure 4. Column (B) shows a three percentage point decline in enrollment within one year. This is consistent with Panel B of Figure 4.

These results suggest that some graduates who would typically have gone to college did not enroll following the time gap. Is this a temporary delay or a permanent decision to forgo further schooling? Columns (C) and (D) measure college enrollment two and three years after the Icfes exam. The point estimates change little relative to column (B). Though the

[146] This yields coefficients identical to those from an individual-level regression but reduces computing time.

Table 2. Switching schools — Time gap and college enrollment
Dependent variable: Enrolled in college within t years of the Icfes

	(A)	(B)	(C)	(D)
Panel A. Benchmark specification				
	0.5 years	1 year	2 years	3 years
Switching schools, 2009 cohort (δ_{hc})	-0.049^{***}	-0.030^{***}	-0.028^{***}	-0.027^{***}
	(0.008)	(0.004)	(0.006)	(0.006)
R^2	0.884	0.904	0.911	0.911
Panel B. High school matching				
	0.5 years	1 year	2 years	3 years
Switching schools, 2009 cohort (δ_{hc})	-0.049^{***}	-0.029^{***}	-0.028^{***}	-0.029^{***}
	(0.007)	(0.004)	(0.006)	(0.006)
R^2	0.890	0.909	0.915	0.915
Panel C. Linear high school trends				
	0.5 years	1 year	2 years	3 years
Switching schools, 2009 cohort (δ_{hc})	-0.032^{**}	-0.023^{***}	-0.023^{**}	-0.019
	(0.015)	(0.008)	(0.009)	(0.011)
R^2	0.915	0.932	0.938	0.939
N	36,972	36,972	36,972	36,972
# regions	33	33	33	33
Dependent var. mean	0.161	0.257	0.338	0.382

Notes: The sample is the 2001–2009 Icfes exam cohorts at high schools in columns (A), (B), (D), and (E) of Table 1. The dependent variables are college enrollment within t years of the Icfes, where t is listed in the column header. All columns report coefficients on the treatment variable, δ_{hc}, which equals one for the 2009 cohort at switching schools (columns (A) and (B) in Table 1). All regressions include high school dummies and cohort dummies. Regressions are at the school-cohort level with observations weighted by the number of exam takers. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 2001–2008 cohorts.

Panel A estimates equation (1). Panel B adds dummies for cells defined by cohort and groups of schools. There are 30 groups based on school ownership (public or private), training (academic/normal, academic & technical, or technical), and Icfes ranking (far superior, superior, high, middle, or low/inferior/far inferior). Each characteristic is time-invariant; a school is private if it is ever private, and I assign each school its highest training level and Icfes ranking. Panel C adds school-specific linear cohort trends to Panel B.

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

data do not extend beyond three years for the 2009 cohort, the lack of catch-up enrollment suggests that the decline in college attendance is not merely transitory.

4.3. **Robustness specifications.** The benchmark results compare switching schools to all high schools in other regions. Table 1 shows that public switching schools are mostly similar to other public schools (columns (A) and (D)), but private schools in the two areas differ significantly (columns (B) and (E)). For example, private school students in other regions have higher test scores and socioeconomic backgrounds. These differences arise because the comparison group includes elite private schools in Bogotá, and because the best private schools in the affected regions stayed on the September calendar (column (C)).

The differences between switching and comparison schools raise concerns if they lead to divergent college enrollment trends. Table 2 depicts two robustness specifications that address this possibility. Panel B alters the implicit comparison group by matching high schools using their characteristics. I define 30 high school groups g based on a school's public/private status, its level of academic training, and its Icfes performance ranking.[147] I then add group-cohort dummies, γ_{gc}, to equation (1). This yields a matching estimator (e.g., Abadie and Imbens, 2002; Imbens, 2004) where matches are based on school traits. The coefficients in Panel B are thus identified only from variation in enrollment outcomes within similar types of high schools. The results are essentially unchanged from Panel A, suggesting that the choice of comparison group does not drive the main results.

Panel C adds linear cohort trends for each high school, $\tilde{\gamma}_h \times c$, to the specification for Panel B. This is the standard differences in differences test of adding linear terms in the "time" dimension (Angrist and Pischke, 2009). The coefficient magnitudes fall slightly relative to previous specifications, which reflects a small divergence in enrollment trends between switching schools and other schools (see Figure 4). Nonetheless, the sharp decrease in enrollment for 2009 switching school graduates is distinguishable from the linear trends, and the pattern of coefficients across years matches that in Panels A and B.

Standard errors in Table 2 are clustered at the region level. Colombia has 33 administrative regions, which is below the rule-of-thumb for inference using the typical cluster-robust standard errors (e.g., Angrist and Pischke, 2009). To address this, Appendix A.5 describes additional regressions that are at the region-cohort level and use a $t(33-2)$ distribution for

[147] See the notes to Table 2 for details on the group definitions. Results are similar with alternative definitions.

inference. This follows Donald and Lang (2007), who recommend "between-group" estimators in settings with few clusters but a large number of observations per cluster (see also Wooldridge, 2003). These region-cohort level regressions yield larger standard errors but do not measurably alter the point estimates or patterns of statistical significance from Table 2.

4.4. **Potential explanations for the enrollment decline.** Table 2 shows that many 2009 switching school graduates experienced an unusual one semester break after high school and did not subsequently enroll in college. One potential concern in attributing the enrollment decline to the time gap is that these students may have been affected by other elements of the calendar transition. For example, the 2009 school year had fewer instructional days prior to the Icfes exam, and more days afterward (see Figure 3). This may have harmed students' exam performance or reduced their likelihood of graduating.

Appendix A.6 tests these hypotheses. The calendar shift did not have a significant impact on the number students who took the Icfes exam at switching schools. There is also no change in students' average Icfes scores. Thus there is little evidence that the transition reduced students' college preparedness.

Another potential hypothesis is that the transition decrease admission rates at selective colleges. This could arise if colleges did not rebalance their January/September cohort slots in response to the shifting calendar. However, Appendix A.7 shows that the enrollment decline occurs mainly at non-selective colleges; thus the primary effect of the transition was a decline in enrollment at colleges with open admissions, not a shift in access to selective colleges.[148]

It is also hard to explain the enrollment decline by a decrease in graduates' returns to college. The time gap reduced post-college careers by one-half year at at most, and the corresponding loss in discounted returns is small relative to estimates of the college earnings premium (Autor, 2014*b*; Zimmerman, 2014).

[148] Further, Section 6 shows that the enrollment decline is largest for low SES and low ability students, who are less likely to attend selective colleges.

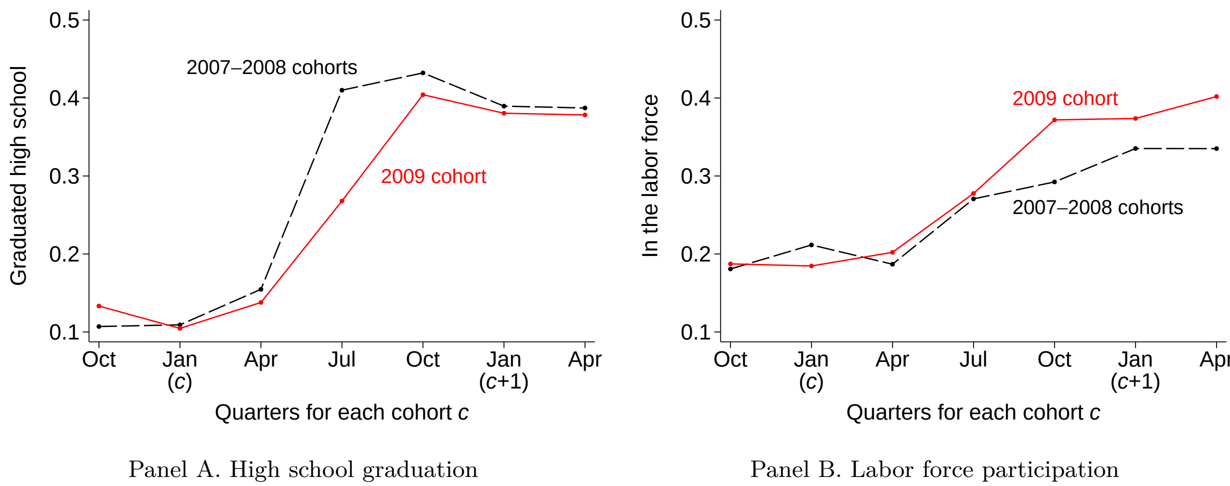


FIGURE 5. 17 year olds in the affected regions

Notes: Data are from the 2007–2010 monthly urban (*cabecera*) and rural (*resto*) GEIH household surveys. The sample is 1990–1992 birth cohorts in the affected regions; this defines cohorts of individuals who turned 17 in 2007–2009. For February 2009 the sample is current 16 year olds because birthdates are missing. The sample includes only children, grandchildren, or other relatives of the household head.

The x-axis combines monthly surveys into three-month quarters. It begins in October before the cohort year and ends in June after the cohort year (e.g., 10/2008 through 6/2010 for the 2009 cohort). 2006 surveys are not available, so Oct–Dec 2006 values for the 2007 cohort are equal to Oct–Dec 2007 values for the 2008 cohort plus the average difference between the 2007 and 2008 cohorts in Jan–Jun of the cohort year.

Panel A shows the fraction of each cohort group with a high school degree or above. Panel B shows the fraction in the labor force, defined as appearing in the employed (*ocupados*) or unemployed (*desocupados*) survey. Panel B omits months with mid-term breaks (1/2007, 1/2008, and 4/2009; see Figure 3) when labor force participation temporarily increases, and it is lagged one month because survey questions refer to labor force activity in the prior 1–4 weeks (e.g., October values are from the November survey).

Calculations use weights that are fixed over time, which are the sum of survey weights across all months within non-missing cells defined by region, age, gender, urban/rural survey, and secondary educated mother.

The time gap more likely influenced college preferences through students' experiences while away from school. Figure 5 provides suggestive evidence of this channel using 2007–2010 Colombian household survey data. The graphs depict two cohorts of individuals in the affected regions: those who turned 17 years old in the pre-transition years (2007–2008), and those who turned 17 in the first year of the calendar transition (2009).

Panel A plots the fraction of individuals in each cohort group with a high school degree. The x-axis displays quarters beginning in October–December before the cohort year; this is the start of the final year of high school for 17 year olds with on-time academic progress.[149]

[149] Age cohort is a noisy measure of graduation cohort because a few students graduate early and many others are behind schedule. This explains why 10 percent of 17 year olds have a high school degree before July and only 40 percent have one after. Figure 5 also cannot separate switching and non-switching schools in the affected regions, but switching school graduates are a large majority of all graduates (see Table 1).

The graduation rate for the 2007–2008 cohorts jumps from 10 to 40 percent in the July quarter but does not reach 40 percent until October for the 2009 cohort. This reflects the delayed calendar for 2009 graduates in many switching schools (see Figure 3).

Panel B depicts labor force participation rates for the same cohorts and time periods. Prior to on-time high school graduation, roughly 20 percent of 17 year olds were either employed or looking for work. These rates increase for all cohorts in July–September as students finish high school. In October–December, however, labor force participation is higher in the 2009 cohort than in the 2007–2008 cohorts. October–December is the gap between high school graduation and potential college enrollment for some 2009 graduates. This result suggests that the time gap led some 2009 graduates to work or look for a job. These experiences may have prompted some students to forgo further schooling, as suggested by the results in Table 2. Consistent with this, Panel B in Figure 5 shows that labor force participation remains higher for the 2009 cohort through the start of the next year.[150]

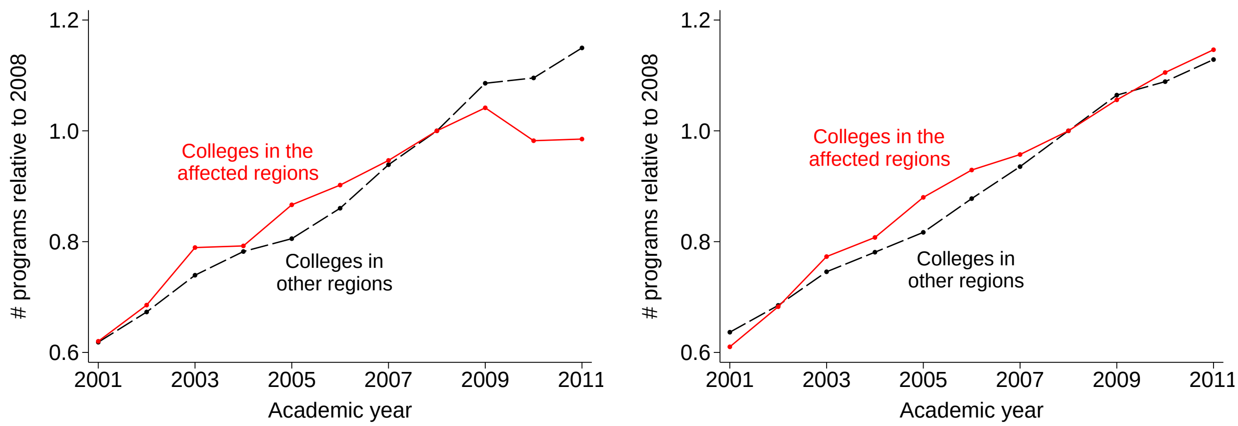
Though the calendar transition had multiple components, Figure 5 suggests the decline in college attendance among 2009 graduates is most consistent with a causal effect of the post-graduation gap. The next section corroborates this finding by analyzing high schools in the affected regions that did not change calendars, which yield an arguably cleaner test of the effect of time gaps on college enrollment.

5. Staying schools

This section shows that the academic calendar shift also affected college entry timing for graduates from schools in the affected regions that did not change calendars, which I call *staying schools*. It then asks how the resulting time gap affected college enrollment.

5.1. **A time gap at staying schools.** Colombian students apply to both a college and a major. Many programs are offered twice per year, while others begin annually. In the affected regions, colleges historically offered their annual programs in the September cohort

[150] Appendix A.8 shows that the findings in Figure 5 are statistically significant in regressions that also include 17 year olds in other regions as a comparison group.



Panel A. September cohort

Panel B. Both cohorts

FIGURE 6. Growth in college programs relative to 2008

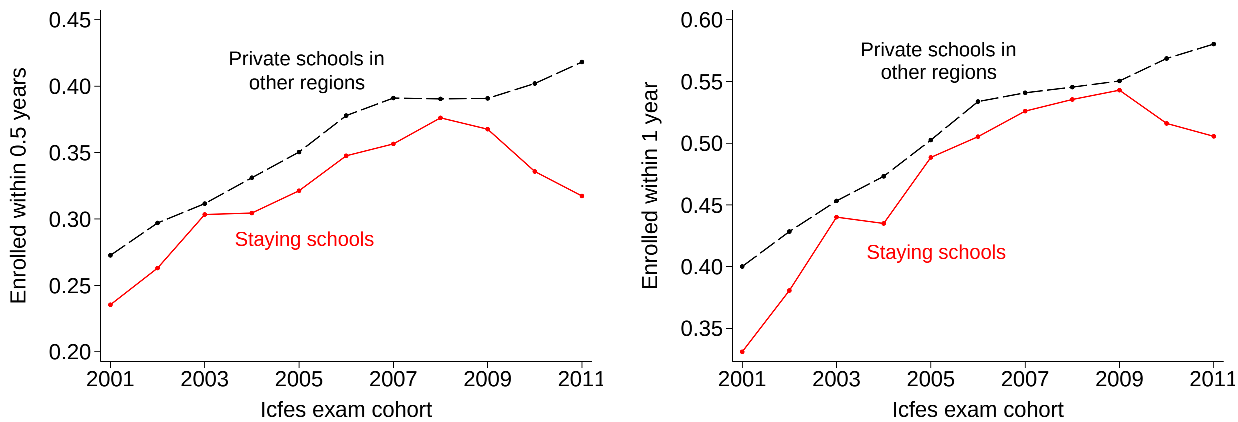
Notes: This graph shows the number of programs with at least one enrollee in each academic year of the Ministry of Education records. Programs are defined by an institution and a non-missing program name. Academic year includes September of the listed year and January of the following year. For both affected regions and other regions, the y-axis is the number of programs in a given year divided by the number of programs in 2008. Panel A shows the number of September programs. Panel B shows the number of September programs plus the number of January programs.

because nearly all local high school students graduated in June. With the calendar shift, the large majority of students now finish in November. As a result, colleges in the affected regions shifted some annual programs to the January cohort.

Figure 6 illustrates this shift in college program timing. Panel A plots the number of programs in the September cohort at colleges in the affected regions and in other regions. The y-axis is normalized to represent the change in programs relative to 2008. In both areas the quantity of September programs increased by more than two-thirds from 2001–2009.[151] In 2010–2011, however, the number of September programs declined by about six percent in the affected regions, falling behind program growth in other regions.

Panel B shows the total number of programs offered across the September and January cohorts. The two areas exhibit similar annual program growth in 2010 and 2011. Figure 6 thus shows that colleges in the affected regions shifted some programs from the September to the January cohort without altering the total annual quantity. This affected the college enrollment timing of the 2010–2011 cohorts at staying schools, which finished in June. Graduates who wanted to apply to programs that moved to the January cohort had to wait an

[151] Program growth reflects new majors at existing colleges and the opening of new colleges.



Panel A. Enrolled within 0.5 years

Panel B. Enrolled within 1 year

FIGURE 7. Staying schools — College enrollment by years since Icfes

Notes: Staying schools include students from column (C) in Table 1. Private schools in other regions include students from column (E). College enrollment is defined as in Table 1.

extra semester to do so. This created another time gap after graduation, induced in this case by changes in college calendars rather than high school calendars.

5.2. **Effects on college enrollment.** I explore the effects of the college program shift by comparing graduates from staying schools (column (C) in Table 1) to graduates from private schools in other regions (column (E)).

Figure 7 shows the college enrollment rates for these two school groups by Icfes exam cohort. Panel A displays the fraction of graduates who began college one semester after taking the Icfes. These initial college entry rates for staying schools and other schools follow similar upward trends through 2009. In 2010–2011, the immediate enrollment rate drops sharply at staying schools. This matches the timing of the college program shift in Figure 6.

Panel A of Figure 7 is a "first stage" result; some graduates could not immediately enter programs that moved to the January cohort. Panel B shows college enrollment rates within one year of the exam, which includes entry into both the September and January cohorts. Since Figure 6 shows no change in annual program offerings, staying school graduates who were waiting for January programs could have enrolled within one year. Panel B suggests that many did not. The enrollment decline in the 2010–2011 cohorts is slightly smaller in magnitude but still evident. This suggests the time gap after high school altered graduates' college attendance decisions.

Table 3 presents the regression version of Figure 7. This table is analogous to Table 2, and the regressions are also based on equation (1). In Table 3, however, the sample is 2001–2011 graduates from staying schools and private high schools in other regions, and the treatment indicator, δ_{hc}, equals one for the 2010–2011 cohorts at staying schools. Since the college enrollment records extend through 2012, this table shows cumulative enrollment at a maximum duration of $t = 1.5$ years after the Icfes.

Panel A shows the benchmark results from equation (1). Columns (A) and (B) match Figure 7. There is a seven percentage point decline in immediate college enrollment for 2010–2011 staying school graduates relative to graduates from other regions. The enrollment decline within one year is slightly smaller at five percentage points.

Column (C) depicts cumulative enrollment within 1.5 years. The magnitude of the coefficient falls slightly relative to column (B), but there is still an enrollment shortfall of four percentage points in staying schools. Data constraints do not allow me to see if this catch-up enrollment continued, but Table 3 shows that many staying school graduates were not enrolled in college one full year after the time gap.[152]

Panels B and C present the same two robustness specifications as in Table 2. Panel B matches treated and comparison high schools by adding dummies for cells defined by cohort and school traits (see Section 4.3). These dummies lower the standard errors but have little effect on the point estimates. Panel C adds high-school-specific linear cohort trends to the specification for Panel B. These terms have little effect on the immediate enrollment coefficient, but they substantially increase the magnitudes of the estimates in columns (B) and (C). This reflects non-parallel enrollment trends in early Icfes cohorts (2001–2005) at staying schools and comparison schools (see Panel B of Figure 7). Nonetheless, these trends

[152] Unlike Table 3, Table 2 shows little evidence of catch-up enrollment at switching schools. This is consistent with evidence in Section 6 that low SES/ability students are more likely to forgo college after the gap.

TABLE 3. Staying schools — Time gap and college enrollment
Dependent variable: Enrolled in college within t years of the Icfes

	(A)	(B)	(C)
Panel A. Benchmark specification			
	0.5 years	1 year	1.5 years
Staying schools, 2010–2011 cohorts (δ_{hc})	-0.070^{***}	-0.050^{***}	-0.039^{***}
	(0.007)	(0.009)	(0.010)
R^2	0.831	0.857	0.866
Panel B. High school matching			
	0.5 years	1 year	1.5 years
Staying schools, 2010–2011 cohorts (δ_{hc})	-0.067^{***}	-0.048^{***}	-0.037^{***}
	(0.005)	(0.007)	(0.008)
R^2	0.839	0.866	0.875
Panel C. Linear high school trends			
	0.5 years	1 year	1.5 years
Staying schools, 2010–2011 cohorts (δ_{hc})	-0.072^{***}	-0.080^{***}	-0.077^{***}
	(0.007)	(0.006)	(0.007)
R^2	0.874	0.899	0.909
N	13,695	13,695	13,695
# regions	25	25	25
Dependent var. mean	0.338	0.482	0.550

Notes: The sample is the 2001–2011 Icfes exam cohorts at high schools in columns (C) and (E) of Table 1. The dependent variables are college enrollment within t years of the Icfes, where t is listed in the column header. All columns report coefficients on the treatment variable, δ_{hc}, which equals one for the 2010–2011 cohorts at staying schools (column (C) in Table 1). All regressions include high school dummies and cohort dummies. Regressions are at the school-cohort level with observations weighted by the number of exam takers. Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 2001–2008 cohorts.

Panel A estimates equation (1). Panel B adds dummies for cells defined by cohort and groups of schools (see Table 2 for details). Panel C adds school-specific linear cohort trends to Panel B.

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

converge in the cohorts prior to the college program shift (2005–2009), and the estimates in Panel C of Table 3 still suggest a large negative impact of the time gap on enrollment.[153]

The results in Table 3 are not attributable to changes in instructional time because staying schools did not alter their academic calendars. This is corroborated by balance tests in Appendix A.6 that show no differential changes in the number of Icfes exam takers or their average scores. Further, the results are not likely due to decreases in college admission rates

[153] Appendix A.5 shows that the results in Table 3 are also robust to the stricter inference procedures based on Donald and Lang (2007) (see Section 4.3).

as there were no alterations in annual program offerings. Instead, these results are evidence of a causal effect of post-high school gaps on college enrollment.

6. Heterogeneous responses to the time gap

Sections 4 and 5 showed that switching school and staying school graduates experienced a time gap after high school, but for different reasons. Both mechanisms, however, led some graduates to forgo further schooling. The net effect was a seven percent decline in college enrollment rates in affected regions.

This section asks whether time gaps can also affect the distribution of schooling outcomes by exploring heterogeneity in these effects at both school types. I first examine whether responses to the gap vary by student characteristics. I then ask if students were more likely to forgo low-paying majors.

6.1. **Heterogeneity by student characteristics.** To study how time gap responses vary across individuals, I estimate coefficients like those in Tables 2 and 3 for different groups of students. However, student types may vary in both baseline college enrollment rates and in their exposure to the time gap. For example, high ability graduates are more likely to enroll in college, and they may also have been more likely to choose college programs that shifted from the September to the January cohorts. This could lead to larger estimates even absent any heterogeneity in responsiveness to the time gap.

To address this, I estimate two stage least squares (2SLS) regressions that relate the long-term enrollment effect to the immediate effect of the transition. Consider the regressions:

$$y_{hc}^{0.5} = \gamma_h + \gamma_c + \beta^{0.5}\delta_{hc} + \epsilon_{hc}, \tag{2}$$

$$y_{hc}^{max} = \alpha_h + \alpha_c + \theta\hat{y}_{hc}^{0.5} + \upsilon_{hc}. \tag{3}$$

Equation (2) is a first stage regression identical to equation (1). The dependent variable, $y_{hc}^{0.5}$, is the fraction of students from high school h and Icfes cohort c who begin college 0.5

years after the exam. The regression includes high school dummies, γ_h, cohort dummies, γ_c, and a treatment indicator for schools and cohorts affected by the calendar shift, δ_{hc}.

Equation (3) is the second stage regression. The dependent variable, y_{hc}^{max}, is cumulative college enrollment at the maximum observable duration since the Icfes exam. The regression also includes school and cohort dummies, but the main independent variable is enrollment within 0.5 years of the exam, $y_{hc}^{0.5}$. The coefficient of interest, θ, measures the effect of a one percentage point change in the immediate college enrollment rate on longer-term enrollment.

College enrollment is endogenous, so the second stage (3) uses predicted values from the first stage (2), $\hat{y}_{hc}^{0.5}$. In other words, the treatment variable, δ_{hc}, is an instrument for $y_{hc}^{0.5}$. The 2SLS coefficient, θ, equals the ratio of the longer-term enrollment effect to the immediate effect ($\beta^{max}/\beta^{0.5}$). Thus θ measures responsiveness to the time gap—the reduction in longer-term enrollment that is attributable to the decline in initial enrollment from the transition.

Column (A) in Table 4 shows the estimate of θ from equation (3). Panel A displays the result for schools that switched to the January calendar. The sample is the same as in Table 2, and the dependent variable is cumulative college enrollment within three years of the Icfes. The coefficient shows that a one percentage point decline in initial enrollment during the transition led to a 0.55 percentage point decrease in the three-year enrollment rate.[154]

Columns (B)–(D) use this 2SLS procedure to compare time gap responsiveness across student types. The columns divide the sample into two groups defined by socioeconomic status, academic ability, and age, respectively.[155] These regressions are similar to column (A), but all terms are interacted with a dummy for one of the groups. In Table 4 this dummy is called "Disadvantaged$_d$" and equals one for: low SES students (column (B)), low ability students (column (C)), and students older than the on-time age (column (D)).[156]

[154] This equals the ratio of the estimates from columns (D) and (A) in Panel A of Table 2.
[155] I do not find heterogeneous responses by gender that are consistent across specifications.
[156] Regressions in columns (B)–(D) are at the school-cohort-Disadvantaged level (hcd); they calculate enrollment rates separately for Advantaged and Disadvantaged students in the same high school and cohort. The full second-stage specification, of which Table 4 displays only the θ and θ^D coefficients, is

$$y_{hcd}^{max} = \alpha_{hd} + \alpha_{cd} + \theta\hat{y}_{hcd}^{0.5} + \theta^D\left(\hat{y}_{hcd}^{0.5} \times \text{Disadvantaged}_d\right) + \upsilon_{hcd}.$$

Table 4. Heterogeneity in responsiveness to the time gap

Panel A. Switching schools
Dependent variable: Enrolled in college within 3 years of the Icfes

	(A)	(B)	(C)	(D)
		Definition of Disadvantaged group		
	All students	Mother no college	Icfes below median	Age 18 or older
Time gap ($\hat{y}^{0.5}_{hcd}$)	0.553***	0.295*	0.385**	0.524***
	(0.165)	(0.174)	(0.177)	(0.183)
Time gap ($\hat{y}^{0.5}_{hcd}$) × Disadvantaged$_d$		0.344***	0.464***	0.250**
		(0.114)	(0.131)	(0.100)
N	36,972	14,672	73,074	73,178
R^2	0.324	0.253	0.261	0.289
# regions	33	33	33	33

Panel B. Staying schools
Dependent variable: Enrolled in college within 1.5 years of the Icfes

		Definition of Disadvantaged group		
	All students	Mother no college	Icfes below median	Age 18 or older
Time gap ($\hat{y}^{0.5}_{hcd}$)	0.564***	0.632***	0.524***	0.449***
	(0.093)	(0.059)	(0.071)	(0.123)
Time gap ($\hat{y}^{0.5}_{hcd}$) × Disadvantaged$_d$		0.282***	0.054	0.235***
		(0.102)	(0.176)	(0.079)
N	13,695	9,583	26,660	26,717
R^2	0.451	0.332	0.397	0.398
# regions	25	25	25	25

Notes: In Panel A, the sample is as in Table 2. The dependent variable is college enrollment within 3 years of the Icfes. The treatment variable, δ_{hc}, equals one for the 2009 cohort at switching schools.

In Panel B, the sample is as in Table 3. The dependent variable is college enrollment within 1.5 years of the Icfes. δ_{hc} equals one for the 2010–2011 cohorts at staying schools.

Column (A) is the 2SLS regression (3), where enrollment within 0.5 years of the Icfes, $y^{0.5}_{hc}$, is instrumented by δ_{hc}. I report only the coefficient on $\hat{y}^{0.5}_{hc}$. The regression includes high school dummies and cohort dummies. The regression is at the school-cohort level with observations weighted by the number of exam takers.

Columns (B)–(D) are the same 2SLS regression, but all terms are interacted with a dummy for the "Disadvantaged" group in the column header. The reported variables, $\hat{y}^{0.5}_{hcd}$ and $\hat{y}^{0.5}_{hcd} \times$ Disadvantaged$_d$, are instrumented by δ_{hc} and $\delta_{hc} \times$ Disadvantaged$_d$. The regression is at the school-cohort-Disadvantaged level with observations weighted by the number of exam takers.

The variables that define Disadvantaged are calculated as in Table 1. Column (B) excludes the 2001–2007 cohorts because mother's education is not available. Median Icfes is defined in the sample for each regression.

Parentheses contain standard errors clustered at the region level.

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

In column (B) of Panel A, the coefficient on $\hat{y}^{0.5}_{hcd}$ is the effect of time gap for switching school graduates whose mothers attended college. The coefficient on the $\hat{y}^{0.5}_{hcd} \times$ Disadvantaged$_d$ interaction is the difference in the time gap effects for students without and with college educated mothers. The effect for graduates without college educated mothers (0.295 + 0.344) is more than double that for high SES students (0.295). In other words, low SES graduates who experienced a time gap after high school were more likely to forgo college.[157]

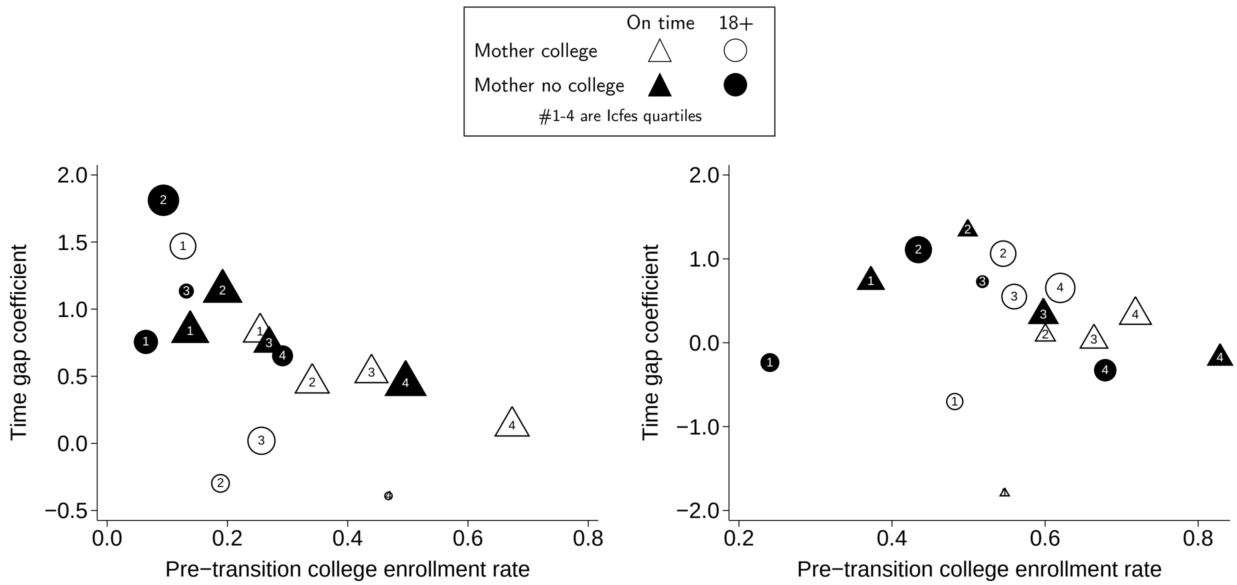
Columns (C) and (D) in Panel A show similar patterns for academic ability and age. Column (C) splits the sample by Icfes scores, and the time gap effect is more than twice as large for graduates with below-median scores. Column (D) compares students with on-time and delayed academic progress; the time gap effect is about 50 percent larger for students who are 18 years old or older.

Panel B of Table 4 shows analogous results for schools in the affected regions that did not change calendars. The sample is the same as in Table 3, and the dependent variable is college enrollment within 1.5 years of the Icfes. Column (A) shows that the mean time gap effect at staying schools is nearly identical to that at switching schools.[158] Columns (B)–(D) show heterogeneity using the same characteristics as in Panel A. The results mirror those for switching schools. "Disadvantaged" staying school graduates are more responsive to the time gap, although the heterogeneity by ability is statistically insignificant.

Table 4 thus shows that low SES, low ability, and older students are more likely to forgo college after the time gap. Another characteristic these groups share is that each has a low overall college enrollment rate. Figure 8 illustrates this using 16 student types defined by mothers' college attendance, age above/below 18, and quartiles of Icfes scores. The y-axis displays time gap coefficients, θ, from separate estimations of equation (3) for each of the 16 types, with larger symbols indicating more precise coefficients. The x-axis depicts the three-year college enrollment rates of each group in the pre-transition cohorts (2001–2008). Panel A displays coefficients from switching schools; Panel B depicts staying school estimates.

The scatterplots in both panels of Figure 8 have a downward shape; student types with low baseline enrollment rates are also more responsive to the time gap.[159] The group definitions

[157] Column (B) includes only the 2008–2011 cohorts for which mother's education is available. Results are similar if I define SES by mean mother's education at the high school level and include all 2001–2011 cohorts.
[158] This coefficient also equals the ratio of the estimates from columns (C) and (A) in Panel A of Table 3.
[159] Some time gap coefficients in Figure 8 are greater than one, suggesting that the gap also reduced the long-term enrollment of students who would not have initially enrolled anyway.



Panel A. Switching schools

Panel B. Staying schools

FIGURE 8. Time gap responsiveness and enrollment rates by SES, ability, and age

Notes: The samples for Panels A and B are the same as in Panels A and B of Table 4, respectively. This figure defines 16 student groups based on the same definitions of SES (two types), ability (four types), and age (two types) as in Table 4, with two exceptions. First, it uses Icfes quartiles rather than the median. Second, mother's education is only available in the 2008–2011 cohorts, so this figure defines SES at the high school level to include all cohorts. Mother no college indicates high schools above the sample median fraction of 2008 graduates without college educated mothers; mother college are high schools below the median.

For each of the 16 student groups, the pre-transition college enrollment rate is the fraction of 2001–2008 affected region graduates who enter college within three years of the Icfes. The time gap coefficient is the estimate of θ from separate regressions (3) for each student group. Regressions are as in column (A) of Table 4. Symbol size is proportional to $\log(1 + 1/\hat{\sigma}_g^2)$, where $\hat{\sigma}_g^2$ is the standard error on $\hat{\theta}_g$ for each group g.

in Figure 8 are analogous to those in Appendix A.1, which shows an inverse relationship between overall enrollment rates and the likelihood of delayed enrollment. Taken together, these results suggest that time gaps may have a doubly negative effect on college enrollment; not only are disadvantaged students more likely to take time off after high school, they are also more likely to forgo college after these time gaps.

A recent literature explores the mechanisms underlying socioeconomic gaps in college attendance (Bailey and Dynarski, 2011; Hoxby and Avery, 2013*b*; Black, Cortes and Lincove, 2015; Clotfelter, Ladd and Vigdor, 2015). Heterogeneity in transition timing may be another factor that contributes to these disparities. The results in Table 4 and Figure 8 are also consistent with a greater influence of time gaps on individuals who are *ex ante* indecisive about college.

6.2. **Heterogeneity by average program earnings.** The previous section shows that time gap effects vary across types of students. This section asks if the time gap also influences students' program choices.

Specifically, I explore whether graduates' post-gap program choices were correlated with their expected earnings in those majors. I calculate program earnings using administrative social security records for 2003–2008 college graduates. Programs are defined by the Ministry of Education's 54 college major groups. For each program, I calculate the average earnings of college graduates with a given Icfes score and gender.[160] This yields a measure of a potential enrollee's predicted earnings in each program based on college graduates with similar characteristics. For each gender and Icfes score combination, I then calculate the median predicted earnings across all 54 programs. I define "high wage" programs as those above the median predicted earnings; "low wage" programs are those below the median.

Table 5 shows regressions similar to those in Tables 2 and 3, but the dependent variables are indicators for enrolling in high or low wage programs. In Panel A, columns (A) and (B) show the effect of the switching school time gap on enrollment in each program type. The time gap led to a similar enrollment decline one semester after the Icfes for both high earnings and low earnings programs. Columns (C) and (D) show the effect on enrollment within three years of the exam. For low wage programs, the three-year enrollment decline is similar to the initial effect. The enrollment effect mostly disappears at high wage programs, and the three-year coefficient is not statistically significant. After the time gap, students were thus more likely to enter programs that yield high earnings, and more likely to forgo programs that deliver low earnings.

[160] Specifically, for each of the 54 programs, I regress log average daily earnings in 2008–2012 on a gender dummy, Icfes percentile, and a quadratic in years since college graduation. Predicted earnings are the sum of the gender and Icfes terms, each multiplied by their coefficients from this regression.

Table 5. Heterogeneity by program earnings
Dependent variable: Enrolled in college within t years of the Icfes

Panel A. Switching schools

	(A)	(B)	(C)	(D)
	Enrolled within 0.5 years		Enrolled within 3 years	
	Low wage programs	High wage programs	Low wage programs	High wage programs
Switching schools, 2009 cohort (δ_{hc})	-0.027^{***}	-0.022^{***}	-0.023^{***}	-0.004
	(0.007)	(0.002)	(0.004)	(0.006)
N	36,972	36,972	36,972	36,972
R^2	0.758	0.845	0.794	0.860
# regions	33	33	33	33
Dependent var. mean	0.076	0.086	0.183	0.199

Panel B. Staying schools

	Enrolled within 0.5 years		Enrolled within 1.5 years	
	Low wage programs	High wage programs	Low wage programs	High wage programs
Staying schools, 2010–2011 cohorts (δ_{hc})	-0.037^{***}	-0.033^{***}	-0.030^{***}	-0.009
	(0.006)	(0.004)	(0.011)	(0.006)
N	13,695	13,695	13,695	13,695
R^2	0.636	0.768	0.665	0.787
# regions	25	25	25	25
Dependent var. mean	0.146	0.192	0.244	0.307

Notes: The samples for Panels A and B are the same as in Tables 2 and 3, respectively. All columns report coefficients on the treatment variable, δ_{hc}, which is defined as in those tables. Regressions are identical to Tables 2 and 3 except for the dependent variables.

The dependent variables measure the fraction of students from each high school and cohort enrolling in high or low wage programs within t years of the Icfes, as listed in the column header. High/low wage programs are defined using social security records for 2003–2008 college graduates, with programs defined by the Ministry of Education's 54 college major groups. For each program (plus one for missing values), I regress log average daily earnings in 2008–2012 on a gender dummy, Icfes percentile, and a quadratic in years since college graduation. I define predicted earnings as the sum of the gender and Icfes terms, each multiplied by their coefficients from this regression. I then calculate the median predicted earnings across all 54 programs for each gender and Icfes score combination in the samples for Panels A and B. "High wage" programs are those above the median predicted earnings; "low wage" programs are those below the median.

Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 2001–2008 cohorts.

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

Panel B shows that this pattern also holds for graduates from staying schools. The coefficients in columns (A) and (B) are similar, suggesting that the initial decline in enrollment is relatively earnings-neutral. Columns (C) and (D) show that the longer-term reduction in enrollment is driven by students forgoing programs with low expected earnings.

Table 5 thus provides suggestive evidence that labor market returns affect students' decisions to return to school after the time gap. This analysis does not study the causal returns to program choice as in Hastings, Neilson and Zimmerman (2015) but rather the descriptive returns, which may be more salient. Related research finds that students' schooling experiences and preferences affect their major choices (Zafar, 2011; Arcidiacono, Hotz and Kang, 2012; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015). My results suggest that time outside of the education system can also shape schooling decisions.

7. Conclusion

An influential literature in labor economics argues that college access lowers wage inequality (Katz and Murphy, 1992*a*; Goldin and Katz, 2008*b*). Recent work shows that one potential barrier to obtaining a college degree is a lack of information. Thus there has been a push by both academics and policymakers to make college costs and outcomes more transparent.

This research and policy agenda focuses on information that is at least potentially attainable prior to the enrollment decision through *search* (Stigler, 1961). For example, Jensen (2010) and Hoxby and Turner (2013*b*) provide students with information on search traits like average returns and tuition, finding substantial effects on their educational choices.

This paper has explored a type of information that is less amenable to information interventions: individuals' *experiences* (Nelson, 1970*b*). In particular, it asked whether breaks after high school graduation affect the decision of whether to enroll in college. To isolate this channel, it exploited a policy that altered academic calendars in some regions of Colombia, which created a one semester gap before potential college entry for some students.

This brief time gap led to a persistent reduction in college enrollment rates in affected regions. The decline was largest for students who were less likely to attend college in the first place. There is also suggestive evidence that labor force participation increased during the time gap, and that individuals who subsequently enrolled chose higher paying majors. These results suggest that students' out-of-school experiences altered their college choices.

The idea that individuals' non-academic experiences matter for their schooling attainment has implications for the design of education systems. Critics of early tracking argue that it forces students into career paths based on limited information from high-pressure exams. Yet such systems may also ease transitions between schooling tiers, which can create educational churning in more flexible systems. Policies that clarify enrollment standards and simplify the application process can also limit academic gaps. The success of initiatives to increase college attainment rates, such as the recent proposal for free community college in the U.S., may depend in part on the degree to which they discourage enrollment delays, with potential implications for wage inequality.

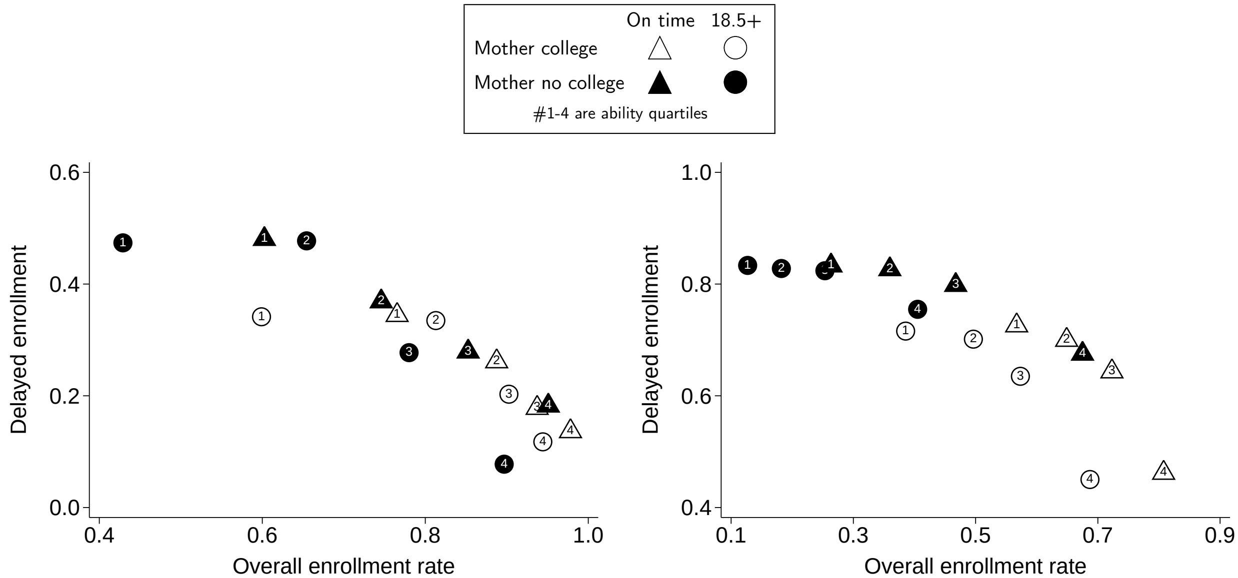
A. Appendix

A.1. **Delayed college enrollment by SES, ability, and age.** Figure 1 in Section 2 shows that delayed enrollment is an important phenomenon in both the U.S. and Colombia. Figure A1 explores the relationship between overall and delayed college enrollment. To do this it divides the samples for Figure 1 into 16 student types based on age, socioeconomic status, and academic ability. For each type, the graph depicts both the overall college enrollment rate (x-axis) and the prevalence of delayed enrollment (y-axis), defined as the fraction of all enrollees who wait more than one semester.

Panel A shows U.S. high school graduates from the NLSY. There is a strong negative relationship between overall and delayed enrollment; students who are less likely to attend college at all are more likely to delay conditional on enrolling. Both delaying and forgoing college are more common among students who are "disadvantaged" in their academic progression, as defined by age, socioeconomic status, and ability.

Panel B shows a similar relationship between delayed and overall enrollment for Colombian high school graduates.

A.2. **Data merging.** The main analysis of this paper uses two administrative datasets: 1) records of Icfes high school exam takers; and 2) records of students enrolled in colleges tracked by the Ministry of Education. I merge these datasets using national ID numbers, birth dates, and names. Nearly all students in both datasets have national ID numbers, but Colombians change ID numbers around age 17. Most students in the Icfes records have the below-17 ID number (*tarjeta*), while the majority of students in the college enrollment 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 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.



Panel A. United States

Panel B. Colombia

FIGURE A1. Delayed college enrollment by SES, ability, and age

Notes: The samples for both panels are students from Figure 1 with non-missing values of the age, SES, and ability variables described below. College enrollment and years since graduation are defined as in Figure 1. Overall enrollment rate is the fraction of graduates who enter college within nine years of graduation. Delayed enrollment is the fraction of all enrollees who began more than six months after graduation.

In Panel A, age is calculated at the end of the month prior to graduation. Mother college means a graduate's mother has some college education. Ability is defined by quartiles of the Armed Services Vocational Aptitude Battery math/verbal percentile within the sample. All calculations use 2011 panel weights.

In Panel B, age is calculated at the end of August in the college entrance exam year. Mother college means a graduate's mother has any degree above basic secondary. Ability is defined by quartiles of the aggregate entrance exam percentile within the sample. See Table 1 for details on these variables.

Merge quality is important because my main dependent variable—enrolling in college—is an indicator for a student's appearance in the enrollment dataset. 41 percent of the 2001–2011 Icfes exam takers appear in the enrollment records, which is broadly comparable to the higher education enrollment rate in Colombia during the same time period.[161] A better indicator of merge success is the percentage of college enrollees that appear in the Icfes records because all domestic college students must take the exam. Among enrollees who took the Icfes exam between 2001 and 2011, I match 91 percent.[162]

[161] The gross tertiary enrollment rate grew from 25 percent to 43 percent between 2001 and 2012 (World Bank World Development Indicators, available at http://data.worldbank.org/country/colombia in December 2015). This rate is not directly comparable to my merge rate because not all high school aged Colombians take the Icfes exam. Roughly 20 percent of the secondary school aged population is not enrolled in high school. This would cause my merge rate to be higher than the World Bank's tertiary enrollment rate.

[162] The enrollment records contain age at time of Icfes for some students, which allows me to calculate the year they took the Icfes exam. Approximately 16 percent of students in the enrollment dataset have missing

Table A1. Higher education institutions in Ministry of Education records

Institution category	(A) # colleges	(B) # exit exam takers/year	(C) % 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.

A.3. **Colleges in the Ministry of Education records.** This section describes the colleges that are included in the Ministry of Education records. For this 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 entrance exam. The exit exam became a requirement for graduation from any higher education institution in 2009.

Column (A) in Table A1 the 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 normative program duration.[163] 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 of exam takers depicted in column (B). Column (C) shows that the Ministry of Education

birth dates, which accounts for the majority of observations I cannot merge. Some duplicate matches arise because students took the Icfes exam more than once, though I erroneously match a small number of students with the same birth date and similar names.

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

records included all Universities but are missing a few colleges that provide more technical training.[164] Overall, 96 percent of exit exam takers attend colleges that appear in the Ministry of Education records.

Another potential issue is that the Ministry of Education's institution coverage has been increasing over time. This could affect the main results if there are differential changes in coverage across regions. Panel B of Figure 6, however, suggests that this is not an issue. This panel depicts the number of college-program pairs that appear in the Ministry of Education records in each academic year. The number of programs is increasing over time due to both program growth and increasing data coverage. But there is no evidence of differential increases between affected and unaffected regions.

In sum, the main results in the paper are likely driven by students forgoing college altogether rather than switching to institutions that are not tracked by the Ministry of Education.

A.4. **Sample of high schools.** This section describes how I select the sample of high schools for my analysis. My sample excludes two categories of schools. First, I exclude high schools that have zero Icfes exam takers in any year between 2001 and 2011. This includes schools for which I cannot cleanly merge in location and academic calendar information.[165] I also drop high schools that are listed in different departments or municipalities over time. This first set of excluded schools includes the 9,548 schools shown in column (A) of Table A2. It also includes six percent of all exam takers who have no high school information.

Second, I exclude high schools that are listed as having a "flexible" academic calendar in any year in 2001–2011. A flexible calendar means that students can begin the school year in either semester. I also omit schools in the affected regions that change calendars before 2010, and schools in other regions that change calendars in any year. These schools were likely

[164] 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*). I also omit one university, *Pontificia Universidad Javeriana*, which has significant variation in the number of enrollees across years in the records. This omission does not affect my main results.

[165] I identify high schools by numeric school IDs, but the Icfes records do not contain these IDs in 2008–2009. I must therefore merge in location and academic calendar data by high school name in these two years. This causes me to drop some high schools that have the same name and time of day (complete day/morning/afternoon) as another school in these years.

TABLE A2. Construction of high school sample

	(A) Unbalanced panel only	(B) Flexible calendar	(C) Final sample
# high schools	9,548	433	4,197
Missing high school	0.06	0.00	0.00
Proportion of all students	0.40	0.05	0.54
# students per school & year	20.1	58.2	61.8
Icfes percentile	0.46	0.41	0.54

Notes: The sample is 11^{th} graders who took the Icfes in 2001–2011. Column (A) includes high schools that are missing exam takers in any year. It also includes schools for which I cannot cleanly merge in location and academic calendar information, schools that are listed in different departments or municipalities over time, and observations with missing school information. Column (B) includes high schools that are listed with a "flexible" academic calendar in any year, affected region schools that change calendars before 2010, and schools in other regions that ever change calendars. Column (C) includes the remaining high schools that have exam takers in every year.

# of students per school & year is the total number of students in 2001–2011 divided by the number of high schools divided by 11. Icfes percentiles are relative to all 11^{th} grade exam takers in the same year and are calculated using the average of the scores from the six core components that did not change in 2001–2011: biology, chemistry, language, mathematics, philosophy, and physics.

more able to adapt to the academic calendar shift. This excludes the 433 schools shown in column (B) of Table A2.

My final sample includes the remaining 4,197 high schools in column (C) of Table A2 (see also Table 1). These schools have Icfes exam takers in every year from 2001–2011, and they contain 54 percent of all high school graduates during this time period. Schools in my sample have 62 students per cohort on average and are larger than excluded schools. Their students also perform better on the Icfes entrance exam.

A.5. **Region-cohort level regressions.** Standard errors in my main analysis are clustered at the region level. Colombia has 33 administrative regions, which is below the rule-of-thumb for inference using the typical cluster-robust standard errors (e.g., Angrist and Pischke, 2009). This section addresses this potential issue by running region-cohort level regressions that mirror the school-cohort level regressions in Tables 2 and 3. This follows Donald and Lang (2007), who recommend "between-group" estimators in settings with few clusters but a large number of observations per cluster (see also Wooldridge, 2003).

I run the differences in differences regression

$$\bar{y}_{rc}^{t} = \gamma_r + \gamma_c + \beta^t \delta_{rc} + \epsilon_{rc}. \tag{A1}$$

This is similar to the benchmark specification (1), but equation (A1) is at the region-cohort level rather than the school-cohort level. The dependent variable, $\bar{y}_{rc}^{t}$, is the mean college enrollment rate t years after the Icfes exam in region r and exam cohort c. The regression includes region dummies, γ_r, cohort dummies, γ_c, and a treatment indicator, δ_{rc}, which equals one for regions and cohorts affected by the transition. Observations are weighted by the number of exam takers. Following the advice in Angrist and Pischke (2009), I use the maximum of OLS and robust standard errors. I use a $t(33-2)$ distribution to calculate statistical significance as suggested by Donald and Lang (2007).

Panel A in Table A3 shows the β^t coefficients from equation (A1) using switching schools and the same comparison group as in Table 2, and with δ_{rc} as an indicator for the 2009 affected region cohort. The coefficients are broadly similar to those in Table 2, but the standard errors are substantially larger in these region-cohort level regressions. Nonetheless, all estimates except for the $t = 3$ coefficient are statistically significant, even with the use of a $t(33-2)$ distribution for inference.

Panels B–D in Table A3 also estimate the region-cohort level regression (A1), but they use dependent variables $\bar{y}_{rc}^{t}$ that more closely reflect the three panels of Table 2. Specifically, I calculate the residuals from three regressions that mirror the panels of Table 2:

$$y_{hc}^{t} = \gamma_h + \epsilon_{hc}, \tag{A2}$$

$$y_{hc}^{t} = \gamma_h + \gamma_{gc} + \epsilon_{hc}, \tag{A3}$$

$$y_{hc}^{t} = \gamma_h + \gamma_{gc} + \tilde{\gamma}_h \times c + \epsilon_{hc}. \tag{A4}$$

γ_h are high school dummies, γ_{gc} are dummies for cells defined by high school groups g and cohort, and $\tilde{\gamma}_h \times c$ are school-specific linear cohort trends (see Table 2 for details).[166] I then

[166] Observations in regressions (A2)–(A4) are also weighted by the number of exam takers.

Table A3. Switching schools — Region-cohort level regressions
Dependent variable: Enrolled in college within t years of the Icfes

	(A)	(B)	(C)	(D)
Panel A. Regional means				
	0.5 years	1 year	2 years	3 years
Affected regions, 2009 cohort (δ_{hc})	-0.046^{***}	-0.026^{**}	-0.023^{*}	-0.022
	(0.009)	(0.011)	(0.013)	(0.013)
R^2	0.905	0.913	0.909	0.914
Panel B. Benchmark specification mean residuals				
	0.5 years	1 year	2 years	3 years
Affected regions, 2009 cohort (δ_{hc})	-0.049^{***}	-0.030^{**}	-0.028^{**}	-0.027^{**}
	(0.009)	(0.011)	(0.012)	(0.012)
R^2	0.782	0.815	0.799	0.800
Panel C. High school matching mean residuals				
	0.5 years	1 year	2 years	3 years
Affected regions, 2009 cohort (δ_{hc})	-0.047^{***}	-0.028^{**}	-0.027^{**}	-0.027^{**}
	(0.009)	(0.012)	(0.012)	(0.013)
R^2	0.093	0.024	0.019	0.019
Panel D. Linear high school trend mean residuals				
	0.5 years	1 year	2 years	3 years
Affected regions, 2009 cohort (δ_{hc})	-0.031^{***}	-0.023^{**}	-0.023^{**}	-0.019^{*}
	(0.010)	(0.011)	(0.010)	(0.010)
R^2	0.044	0.018	0.020	0.015
N	297	297	297	297
# regions	33	33	33	33

Notes: The sample is as in Table 2. All columns report coefficients on the treatment variable, δ_{rc}, which equals one for the 2009 cohort in the affected regions. All regressions include region dummies and cohort dummies. Regressions are at the region-cohort level with observations weighted by the number of exam takers. Parentheses contain the maximum of OLS and robust standard errors.

In Panel A, the dependent variables are mean region-cohort level college enrollment within t years of the Icfes, where t is listed in the column header. The dependent variables in Panels B, C, and D are the mean region-cohort residuals from equations (A2), (A3), (A4), respectively. See Table 2 for details.

Inference uses a $t(33-2)$ distribution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

calculate the mean of the residuals from these regressions for each region r and cohort c. The region-cohort mean residuals from (A2)–(A4) are the dependent variables $\bar{y}^t_{rc}$ in Panels B–D of Table A3. The regressions and inference procedures are otherwise identical to those for Panel A.

The coefficients in Panels B–D of Table A3 closely correspond to those in Table 2, but the standard errors are in most cases substantially larger. However, all estimates are still statistically significant at the ten percent level.

Table A4 repeats these region-cohort level regressions for staying schools and the same comparison group as in Table 3. The table and methods are otherwise identical to those in Table A3 except the treatment variable, δ_{rc}, equals one for the 2010–2011 cohorts in the affected regions. The comparison group for these regressions excludes public schools, and private schools appear in only 25 regions in my sample. Thus inference in Table A4 is based on a $t(25-2)$ distribution.

The results in Table A4 mirror those in Table 3. The point estimates are similar in both tables, and despite larger standard errors in the region-cohort level regressions, nearly all coefficients are statistically significant.

In sum, the results in Tables A3 and A4 show that the main results of this paper are robust to stricter inference methods that address the relatively small number of clusters.

A.6. **Balance tests.** This section tests for effects of the calendar transition on the number of Icfes exam takers or their scores. This is a potential concern at switching schools, where students may have been affected by other elements of the transition such as changes in instructional time. I also test for effects at staying schools, although other transition elements are less of a concern as these schools did not change their academic calendars.

Panel A of Table A5 shows the results of these balance tests for switching schools. The sample is the same as in Table 2. I use the same benchmark regression (1) with different dependent variables. In column (A), the dependent variable is the number of Icfes exam takers in each high school and cohort. The differences in differences coefficient shows that the number of exam takers declined by two students in the 2009 cohort at switching schools relative to schools in other regions. This change is small relative to the mean of 62 students per school-cohort and is not statistically significant.

TABLE A4. Staying schools — Region-cohort level regressions
Dependent variable: Enrolled in college within t years of the Icfes

	(A)	(B)	(C)
Panel A. Regional means			
	0.5 years	1 year	1.5 years
Affected regions, 2010–2011 cohorts (δ_{hc})	-0.060^{***}	-0.039^{**}	-0.028
	(0.014)	(0.016)	(0.017)
R^2	0.932	0.931	0.928
Panel B. Benchmark specification mean residuals			
	0.5 years	1 year	1.5 years
Affected regions, 2010–2011 cohorts (δ_{hc})	-0.069^{***}	-0.050^{***}	-0.039^{**}
	(0.015)	(0.016)	(0.016)
R^2	0.847	0.865	0.847
Panel C. High school matching mean residuals			
	0.5 years	1 year	1.5 years
Affected regions, 2010–2011 cohorts (δ_{hc})	-0.072^{***}	-0.053^{***}	-0.042^{**}
	(0.014)	(0.016)	(0.016)
R^2	0.164	0.128	0.112
Panel D. Linear high school trend mean residuals			
	0.5 years	1 year	1.5 years
Affected regions, 2010–2011 cohorts (δ_{hc})	-0.074^{***}	-0.082^{***}	-0.079^{***}
	(0.015)	(0.014)	(0.014)
R^2	0.106	0.133	0.128
N	275	275	275
# regions	25	25	25

Notes: The sample is as in Table 3. All columns report coefficients on the treatment variable, δ_{hc}, which equals one for the 2010–2011 cohorts in the affected regions. All regressions include region dummies and cohort dummies. Regressions are at the region-cohort level with observations weighted by the number of exam takers. Parentheses contain the maximum of OLS and robust standard errors.

In Panel A, the dependent variables are mean region-cohort level college enrollment within t years of the Icfes, where t is listed in the column header. The dependent variables in Panels B, C, and D are the mean region-cohort residuals from equations (A2), (A3), (A4), respectively. See Table 3 for details.

Inference uses a $t(25-2)$ distribution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (B) tests for effects of the transition on students' entrance exam scores by using Icfes percentile as the dependent variable. The point estimate of the transition effect is one-half of one percentile in magnitude and is not statistically significant.

Table A5. Balance tests

Panel A. Switching schools

	(A)	(B)
	Dependent variable	
	# exam takers	Icfes percentile
Switching schools, 2009 cohort (δ_{hc})	-2.281	-0.005
	(1.816)	(0.009)
N	36,972	36,972
R^2	0.885	0.889
# regions	33	33
Dependent var. mean	61.537	0.527

Panel B. Staying schools

	(A)	(B)
	Dependent variable	
	# exam takers	Icfes percentile
Staying schools, 2010–2011 cohorts (δ_{hc})	-1.696	-0.002
	(1.832)	(0.004)
N	13,695	13,695
R^2	0.892	0.913
# regions	25	25
Dependent var. mean	48.277	0.664

Notes: The samples for Panels A and B are the same as in Tables 2 and 3, respectively. All columns report coefficients on the treatment variable, δ_{hc}, which is defined as in those tables. Regressions are at the school-cohort level with observations unweighted in column (A) and weighted by the number of exam takers in column (B). Regressions are otherwise identical to Tables 2 and 3 except for the dependent variables.

In column (A), the dependent variable is the number of Icfes exam takers in each high school and cohort. The dependent variable in column (B) is the school-cohort mean Icfes percentile (see Table 1).

Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 2001–2008 cohorts.

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

Panel B replicates these regressions for staying schools. The sample and regressions are identical to those in Table 3. In both columns, the point estimates are smaller than those in Panel A and are insignificant. There is no evidence of differential changes in the number of exam takers or their performance in the 2010–2011 cohorts at staying schools.

Thus, Table A5 suggests that the decline in college enrollment at switching schools and staying schools is not driven by changes in the student population or their college preparation.

A.7. **Heterogeneity by college selectivity.** This section explores the hypothesis that the observed college enrollment decline for switching school graduates was driven by an increased difficulty in gaining admission to selective colleges. This could arise if affected region colleges did not rebalance their January/September cohort slots in response to the shifting academic calendar.

To test this hypothesis, I run regressions that are similar to the benchmark specification in Table 2, but I use dependent variables that differ according to the selectivity of the college of enrollment. I use aggregate Ministry of Education data on the number of applicants and admitted students to calculate the admission rate for each college in the years before the calendar transition (2007–2008). I define "less selective" colleges as those above the median admission rate, and "more selective" colleges as those below the median. I then run regressions as in Table 2 with dependent variables that measure only enrollment in less selective or more selective colleges.

Panel A in Table A6 shows the results for switching school graduates using the same sample as in Table 2. Columns (A) and (B) show the enrollment effect one semester after the Icfes exam. The decline in enrollment at switching schools occurs at both less selective and more selective colleges with a similar magnitude. Columns (C) and (D) show the effect on enrollment within three years of the exam. For less selective colleges, the three-year enrollment decline is similar to the initial effect. The enrollment decline mostly disappears at selective colleges and is not statistically significant.

Panel A thus shows that the results in Table 2 are driven by students forgoing colleges with open enrollment and not by an enrollment decline at selective colleges. This argues against the hypothesis that changes in admission rates are driving the main results. The results in Panel A of Table A6 are also consistent with evidence in Section 6 showing that low SES and low ability students were more likely to forgo college during the transition, as these students less frequently attend selective colleges.

Panel B of Table A6 replicates the regressions in Panel B for the staying school sample from Table 3. Unlike the results for switching schools, both the initial and longer-term

Table A6. Heterogeneity by college selectivity
Dependent variable: Enrolled in college within t years of the Icfes

Panel A. Switching schools

	(A)	(B)	(C)	(D)
	Enrolled within 0.5 years		Enrolled within 3 years	
	Less selective colleges	More selective colleges	Less selective colleges	More selective colleges
Switching schools, 2009 cohort (δ_{hc})	-0.024^{***}	-0.025^{**}	-0.020^{*}	-0.007
	(0.004)	(0.010)	(0.010)	(0.007)
N	36,972	36,972	36,972	36,972
R^2	0.816	0.829	0.854	0.857
# regions	33	33	33	33
Dependent var. mean	0.080	0.081	0.186	0.196

Panel B. Staying schools

	(A)	(B)	(C)	(D)
	Enrolled within 0.5 years		Enrolled within 1.5 years	
	Less selective colleges	More selective colleges	Less selective colleges	More selective colleges
Staying schools, 2010–2011 cohorts (δ_{hc})	-0.025^{***}	-0.044^{***}	0.009	-0.048^{***}
	(0.007)	(0.004)	(0.012)	(0.006)
N	13,695	13,695	13,695	13,695
R^2	0.713	0.856	0.760	0.875
# regions	25	25	25	25
Dependent var. mean	0.154	0.183	0.258	0.292

Notes: The samples for Panels A and B are the same as in Tables 2 and 3, respectively. All columns report coefficients on the treatment variable, δ_{hc}, which is defined as in those tables. Regressions are identical to Tables 2 and 3 except for the dependent variables.

The dependent variables measure the fraction of students from each high school and cohort enrolling in less or more selective colleges within t years of the Icfes, as listed in the column header. College selectivity is defined using aggregate application data from the Ministry of Education (available in December 2015 at http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212400.html). The sample for this calculation includes technical- and university-level programs with a non-zero number of applicants and admitted students in the pre-transition periods for which data are available (2007–2008). I calculate the admission rate for each college by dividing the total number of admitted students by the total number of applicants over this time period. I then calculate the median admission rate across all colleges in the samples for Panels A and B. "Less selective" colleges are those above the median admission rate or that do not appear in the aggregate application data; "more selective" colleges are those below the median.

Parentheses contain standard errors clustered at the region level. Dependent variable means are calculated from the 2001–2008 cohorts.

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

enrollment declines at staying schools are driven primarily by more selective colleges. This reflects the fact that time gap for staying school graduates was due to a shift in annual programs from the September to the January calendar (see Figure 6). Annual programs typically cover specialized subjects and are thus offered primarily by selective colleges; non-selective colleges mainly offer semiannual programs in only the most popular subjects. Thus one would expect the staying school enrollment decline to occur primarily at more selective colleges that shifted their academic calendars, consistent with Panel B.

A.8. **Colombian household survey regressions.** Figure 5 in Section 4 provides suggestive evidence that many 2009 switching school graduates joined the labor force during the time gap created by the calendar transition. This section explores the statistical significance of this finding in a regression framework.

For this I use the same 2007–2010 waves of the Colombian household survey as in Figure 5. As in that figure, I define three cohorts of individuals denoted by c: those who turned 17 years old in the pre-transition years ($c \in$ 2007–2008), and those who turned 17 in the first year of the calendar transition ($c = 2009$).

Figure 5 plots dependent variables against time since the start of each cohort's last year of high school. In this section I use t to denote months with $t = 0$ representing the beginning of the last high school grade for students with on-time progression. The regressions below include the first seven quarters following the start of this last grade ($t \in$ 1–21). Further, while Figure 5 includes only 17 year olds in the affected regions, below I also include individuals in other regions as an additional comparison group. I use r to denote regions.

I estimate the regression

$$\bar{y}_{rct} = \gamma_{rc} + \gamma_{rt} + \gamma_{ct} + \beta_p \delta_{rc} + \epsilon_{rct}. \tag{A5}$$

The dependent variable, $\bar{y}_{rct}$, is a mean outcome in a region-cohort-month cell. The regression includes region-cohort dummies, γ_{rc}, region-month dummies, γ_{rt}, and cohort-month dummies, γ_{ct}. The treatment indicator, δ_{rc}, equals one for the 2009 cohort in the affected regions. I allow the treatment coefficient, β_p, to vary with time periods p that capture the quarter of typical high school graduation ($t \in$ 10–12), the time gap quarter for the 2009 cohort ($t \in$ 13–15), and the period following the time gap ($t \in$ 16–21).

Equation (A5) is thus a triple differences regression. The "time" dimension is months since the start of the last year of high school, t, and it uses both age cohorts c and regions r as comparison groups. Treatment is defined for the 2009 cohort in the affected regions as the months of delayed graduation and the periods following it ($t \geq 10$). β_p measures how the change in the outcome from previous months compares to analogous changes in other regions and cohorts. I weight observations by the sum of survey weights in each region-cohort-month cell. Since equation (A5) is a "between-groups" regression, I use the same inference procedures as in Appendix A.5.[167]

Table A7 shows the results from regression (A5). The dependent variable in column (A) is the fraction of individuals with a high school degree. The table reports the three β_p coefficients from this regression, which corresponds to the time periods listed in the beginning of each row for the 2009 cohort in the affected regions. July–September is the quarter in which affected region students typically finished high school in the years before the calendar transition. The first coefficient in Table A7 shows that in July–September 2009, the fraction of affected region 17 year olds reporting a high school degree falls by 15 percentage points relative to other cohorts and regions. This graduation effect disappears by the last quarter of 2009, as the other two coefficients in column (A) are near zero. Column (A) thus replicates the finding from Panel A in Figure 5 that the academic calendar transition led to a one-quarter graduation delay for some students in the 2009 cohort.

Column (B) of Table A7 also estimates equation (A5), but the dependent variable is an indicator for labor force participation. I define individuals as in the labor force if they report being either employed or looking for work in the household survey data. The first coefficient shows that there is no differential change in labor force participation in July–September 2009. However, affected region labor force participation rates increase by 11 percentage points in October–December 2009 relative to other regions and cohorts. This is the time gap between high school graduation and potential college enrollment for some 2009 graduates in

[167] Specifically, I use the maximum of OLS and robust standard errors (Angrist and Pischke, 2009), and I use a $t(24-4)$ distribution for inference (Donald and Lang, 2007) given 24 regions in the survey data and three coefficients of interest.

Table A7. 17 year olds in Colombian household survey data

	(A)	(B)
	Dependent variable	
Affected regions, 2009 cohort (δ_{rc}) in ...	Graduated high school	In the labor force
July−September 2009	-0.149^{***}	-0.011
	(0.036)	(0.036)
October−December 2009	-0.054	0.109^{*}
	(0.057)	(0.053)
January−June 2010	-0.037	0.057^{*}
	(0.036)	(0.029)
N	1,506	1,504
R^2	0.868	0.816
# regions	24	24
Dependent var. mean	0.372	0.244

Notes: Data are from the 2007–2010 monthly urban (*cabecera*) and rural (*resto*) GEIH household surveys. The sample is the 1990–1992 birth cohorts; this defines cohorts of individuals who turned 17 in 2007–2009. For February 2009 the sample is current 16 year olds because birthdates are missing. The sample includes only children, grandchildren, or other relatives of the household head.

The columns estimate the region-cohort-month level regression (A5) with t defined as months since the start of 11^{th} grade for most high schools in the region. This means that $t = 0$ in September before the cohort year in the affected regions, and $t = 0$ in February of the cohort year in other regions. The sample includes months $t = 1$ to $t = 21$ for each cohort. 2006 surveys are not available, so values at $t \leq 3$ are missing for the 2007 cohort in the affected regions.

The dependent variable in column (A) is the fraction of each region-cohort-month cell with a high school degree or above. In column (B), the dependent variable is the fraction of each region-cohort-month cell in the labor force, defined as appearing in the employed (*ocupados*) or unemployed (*desocupados*) survey. This variable is lagged one month because survey questions refer to labor force activity in the prior 1–4 weeks (e.g., October values are from the November survey).

All columns report coefficients β_p on the treatment variable, δ_{rc}, which equals one for the 2009 cohort in the affected regions. This variable is interacted with dummies for three time periods p defined by $t \in$ 10–12, $t \in$ 13–15, and $t \in$ 16–21. This corresponds to the time periods listed in the beginning of each row for the 2009 cohort in the affected regions.

All regressions include region-cohort dummies, region-month dummies, and cohort-month dummies. Regressions use weights that are fixed over time, which are the sum of survey weights across all months within non-missing cells defined by region, age, gender, urban/rural survey, and secondary educated mother. Parentheses contain the maximum of OLS and robust standard errors. Dependent variable means are calculated from the 2007–2008 cohorts.

Inference uses a $t(24-4)$ distribution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the affected regions. Like Panel B of Figure 5, this result suggests that the time gap led some 2009 graduates to work or look for a job, but Table A7 shows that this effect is statistically significant at the ten percent level.

The last coefficient in column (B) shows that the labor force participation increase for the 2009 cohort persists into the first half of 2010 and is statistically significant. The magnitude of this effect falls by roughly 50 percent, consistent with the finding in Table 2 that the initial effect of the time gap on college enrollment declines by one-half in subsequent periods.

In sum, Table A7 shows that the effects of the calendar transition on high school graduation and labor force participation documented in Figure 5 are statistically significant in a regression framework.

References

Abadie, Alberto, and Guido Imbens. 2002. "Simple and Bias-corrected Matching Estimators for Average Treatment Effects." National Bureau of Economic Research.

Abdulkadiroglu, Atila., and T. Sonmez. 2013. "Matching Markets: Theory and Practice." In *Advances in Economics and Econometrics*. Vol. Vol I: Economic Theory, , ed. D Acemoglu, M Arellano and E Dekel, Chapter 1, 3–47. Cambridge University Press.

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica*, 82(1): 137–196.

Abdulkadiroğlu, Atila, Parag Pathak, and Alvin Roth. 2005. "The New York City High School Match." *American Economic Review*, 95(2): 364–367.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." *Handbook of Labor Economics*, 4: 1043–1171.

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

Angrist, Joshua, and Victor Lavy. 2009. "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial." *American Economic Review*, 99(4): 1384–1414.

Angrist, Joshua D, and Alan B Krueger. 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association*, 87(418): 328–336.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics*, 1(1): 136–163.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review.*

Antonovics, Kate, and Ben Backes. 2014. "The Effect of Banning Affirmative Action on College Admissions Policies and Student Quality." *Journal of Human Resources*, 49(2): 295–322.

Arcidiacono, Peter. 2004. “Ability Sorting and the Returns to College Major.” *Journal of Econometrics*, 121(1): 343–375.

Arcidiacono, Peter, and Michael Lovenheim. 2016. “Affirmative Action and the Quality-Fit Tradeoff.” *Journal of Economic Literature*, 54(1): 3–51.

Arcidiacono, Peter, Esteban Aucejo, Patrick Coate, and V Joseph Hotz. 2014. “Affirmative Action and University Fit: Evidence from Proposition 209.” *IZA Journal of Labor Economics*, 3(1): 1.

Arcidiacono, Peter, Esteban M Aucejo, and V Joseph Hotz. 2016. “University Differences in the Graduation of Minorities in STEM Fields: Evidence from California.” *American Economic Review*, 106(3): 525–562. National Bureau of Economic Research.

Arcidiacono, Peter, Esteban M Aucejo, Hanming Fang, and Kenneth I Spenner. 2011. “Does Affirmative Action Lead to Mismatch? A New Test and Evidence.” *Quantitative Economics*, 2(3): 303–333.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo. 2010. “Beyond Signaling and Human Capital: Education and the Revelation of Ability.” *American Economic Journal: Applied Economics*, 2(4): 76–104.

Arcidiacono, Peter, V Joseph Hotz, and Songman Kang. 2012. “Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals.” *Journal of Econometrics*, 166(1): 3–16.

Atkin, David. 2012. “Endogenous Skill Acquisition and Export Manufacturing in Mexico.” National Bureau of Economic Research.

Autor, David H. 2014*a*. “Skills, education, and the rise of earnings inequality among the ’other 99 percent’.” *Science*, 344(6186): 843–851.

Autor, David H. 2014*b*. “Skills, Education, and the Rise of Earnings Inequality Among the ‘Other 99 percent’.” *Science*, 344(6186): 843–851.

Backes, Ben. 2012. “Do Affirmative Action Bans Lower Minority College Enrollment and Attainment? Evidence from Statewide Bans.” *Journal of Human Resources*, 47(2): 435–455.

Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. 2016. “Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India.” *American Economic Review*, 106(6): 1495–1521.

Bailey, Martha J, and Susan M Dynarski. 2011. “Gains and Gaps: Changing Inequality in U.S. College Entry and Completion.” National Bureau of Economic Research.

Becker, Gary S. 1973. “A Theory of Marriage: Part I.” *Journal of Political Economy*, 81(4): 813–846.

Bedard, Kelly, and Elizabeth Dhuey. 2006. “The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects.” *The Quarterly Journal of Economics*, 121(4): 1437–1472.

Beffy, Magali, Denis Fougere, and Arnaud Maurel. 2012. “Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter?” *Review of Economics and Statistics*, 94(1): 334–347.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. 2010. “Affirmative Action in Education: Evidence from Engineering College Admissions in India.” *Journal of Public Economics*, 94(1): 16–29.

Bettinger, Eric P, and Bridget Terry Long. 2009. “Addressing the Needs of Underprepared Students in Higher Education Does College Remediation Work?” *Journal of Human Resources*, 44(3): 736–771.

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. “The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment.” *The Quarterly Journal of Economics*, 127(3): 1205–1242.

Bhattacharya, Debopam. 2009. “Inferring Optimal Peer Assignment from Experimental Data.” *Journal of the American Statistical Association*, 104(486): 486–500.

Bishop, John. 2004. “Drinking from the fountain of knowledge: Student incentives to study and learn–externalities, information problems, and peer pressure.” Center for Advanced Human Resource Studies Working paper.

Black, D. A., and J. A. Smith. 2006. “Estimating the returns to college quality with multiple proxies for quality.” *Journal of Labor Economics*, 24(3): 701–728.

Black, Dan A, Terra G McKinnish, and Seth G Sanders. 2005. “Tight Labor Markets and the Demand for Education: Evidence from the Coal Boom and Bust.” *Industrial and Labor Relations Review*, 59(1): 3–16.

Black, Sandra E, Kalena E Cortes, and Jane Arnold Lincove. 2015. “Apply Yourself: Racial and Ethnic Differences in College Application.” National Bureau of Economic Research.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2011. “Too Young to Leave the Nest? The Effects of School Starting Age.” *The Review of Economics and Statistics*, 93(2): 455–467.

Booij, Adam S, Edwin Leuven, and Hessel Oosterbeek. 2015. “Ability Peer Effects in University: Evidence from a Randomized Experiment." IZA Discussion Paper No. 8769.

Bound, John, Brad Herschbein, and Bridget Terry Long. 2009. "Playing the Admissions Game: Student Reactions to Increasing College Competition." *Journal of Economic Perspectives*, 23(4): 119–146.

Burdett, Kenneth. 1978. "A theory of employee job search and quit rates." *The American Economic Review*, 68(1): 212–220.

Cabral, Luis, and Ali Hortacsu. 2010. "The Dynamics Of Seller Reputation: Evidence From Ebay." *The Journal of Industrial Economics*, 58(1): 54–78.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. "Bootstrap-based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3): 414–427.

Canaan, Serena, and Pierre Mouganie. 2015. "Returns to Education Quality for Low-Skilled Students: Evidence from a Discontinuity." Working Paper. Available at SSRN 2518067.

Card, David. 1995. "Using geographic variation in college proximity to estimate the return to schooling." In *Aspects of labor market behaviour: Essays in Honour of John Vanderkamp.* , ed. E.K. Christofides and R. Swidinsky. Toronto, Ontario:University of Toronto Press.

Card, David, and Alan B Krueger. 1996. "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina." *The Journal of Economic Perspectives*, 10(4): 31–50.

Card, David, and Alan Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *The Journal of Political Economy*, 100(1): 1–40.

Carrell, Scott E, Bruce I Sacerdote, and James E West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica*, 81(3): 855–882.

Castleman, Benjamin L, and Lindsay C Page. 2014. "A Trickle or a Torrent? Understanding the Extent of Summer "Melt" Among College-Intending High School Graduates." *Social Science Quarterly*, 95(1): 202–220.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J Notowidigdo. 2015. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." National Bureau of Economic Research.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American*

Economic Review, 104(9): 2633–2679.

Chiappori, Pierre-Andre, and Bernard Salanie. 2016. "The Econometrics of Matching Models." *Journal of Economic Literature*, 54(3): 832–861.

Choo, Eugene, and Aloysius Siow. 2006*a*. "Estimating a Marriage Matching Model with Spillover Effects." *Demography*, 43(3): 463–490.

Choo, Eugene, and Aloysius Siow. 2006*b*. "Who Marries Whom and Why." *Journal of Political Economy*, 114(1): 175–201.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor. 2015. "Public Universities, Equal Opportunity, and the Legacy of Jim Crow: Evidence from North Carolina." National Bureau of Economic Research.

Cortes, Kalena 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.

Crawford, Vincent P, and Elsie Marie Knoer. 1981. "Job Matching with Heterogeneous Firms and Workers." *Econometrica*, 49(2): 437–450.

Cullen, Julie Berry, Mark C Long, and Randall Reback. 2013. "Jockeying for Position: Strategic High School Choice under Texas' Top Ten Percent Plan." *Journal of Public Economics*, 97: 32–48.

Dale, Stacy B, and Alan B Krueger. 2014*a*. "Estimating the Effects of College Characteristics Over the Career using Administrative Earnings Data." *Journal of Human Resources*, 49(2): 323–358.

Dale, Stacy Berg, and Alan B Krueger. 2002*a*. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *The Quarterly Journal of Economics*, 117(4): 1491–1527.

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

Dale, Stacy Berg, and Alan B. Krueger. 2014*b*. "Estimating the effects of college characteristics over the career using administrative earnings data." *The Journal of Human Resources*, 49(2): 323–358.

Daugherty, Lindsay, Paco Martorell, and Isaac McFarlin. 2014. "Percent Plans, Automatic Admissions, and College Outcomes." *IZA Journal of Labor Economics*, 3(1): 1.

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2016. "The Determinants of Mismatch Between Students and Colleges." *Journal of Labor Economics* (forthcoming).

Dinkelman, Taryn, and Claudia Martínez A. 2014. "Investing in Schooling in Chile: The Role of Information about Financial Aid for Higher Education." *Review of Economics and Statistics*, 96(2): 244–257.

Dobkin, Carlos, and Fernando Ferreira. 2010. "Do School Entry Laws affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review*, 29(1): 40–54.

Donald, Stephen G, and Kevin Lang. 2007. "Inference with Difference-in-differences and Other Panel Data." *The Review of Economics and Statistics*, 89(2): 221–233.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, 48(4): 935–963.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*, 101(5): 1739–1774.

Durlauf, Steven N. 2008. "Affirmative Action, Meritocracy, and Efficiency." *Politics, Philosophy & Economics*, 7(2): 131–158.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker. 2016. "Referral-based job search networks." *The Review of Economic Studies*, 83(2): 514–546.

Dynarski, Susan, and Judith Scott-Clayton. 2013. "Financial Aid Policy: Lessons from Research." *The Future of Children*, 23(1): 67–91.

Epple, Dennis, and Richard E. Romano. 1998*a*. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review*, 88(1): 33–62.

Epple, Dennis, and Richard Romano. 1998*b*. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review*, 88(1): 33–62.

Epple, D., R. Romano, and H. Sieg. 2006. "Admission, tuition, and financial aid policies in the market for higher education." *Econometrica*, 74(4): 885–928.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics*, 111(4): 1007–47.

Figlio, David, and Susanna Loeb. 2011. "School Accountability." Vol. 3, 383–421.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. 2016. "Cultural Proximity and Loan Outcomes." *American Economic Review* (forthcoming).

Fredriksson, Peter, and Björn Öckert. 2014. "Life-cycle Effects of Age at School Start." *The Economic Journal*, 124(579): 977–1004.

Fryer, Roland G. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *The Quarterly Journal of Economics*, 126(4): 1755–1798.

Gale, David, and Lloyd S Shapley. 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly*, 69(1): 9–15.

Gibbons, Robert, and Lawrence F. Katz. 1991. "Layoffs and Lemons." *Journal of Labor Economics*, 9(4): 351–80.

Goldin, Claudia, and Lawrence F. Katz. 2008*a*. *The race between education and technology.* Cambridge, MA:Belknap Press of Harvard University Press.

Goldin, Claudia Dale, and Lawrence F Katz. 2008*b*. *The Race Between Education and Technology.* Harvard University Press.

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith. 2014. "College Access, Initial College Choice and Degree Completion." NBER Working Paper 20996.

Goodman, Sarena F. 2013. "Learning from the Test: Raising Selective College Enrollment by Providing Information." Board of Governors of the Federal Reserve System.

Graham, Bryan. 2011. "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers." *Handbook of Social Economics*, 1: 965–1052.

Graham, Bryan S. 2008. "Identifying Social Interactions through Conditional Variance Restrictions." *Econometrica*, 76(3): 643–660.

Graham, Bryan S, Guido W Imbens, and Geert Ridder. 2010. "Measuring the Effects of Segregation in the Presence of Social Spillovers: A Nonparametric Approach." National Bureau of Economic Research.

Graham, Bryan S, Guido W Imbens, and Geert Ridder. 2014. "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis." *Quantitative Economics*, 5(1): 29–66.

Graham, Bryan S, Guido W Imbens, and Geert Ridder. 2016. "Identification and Efficiency Bounds for the Average Match Function under Conditionally Exogenous Matching." National Bureau of Economic Research Working Paper No. 22098.

Hanushek, Eric A., and Ludger Woessmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries." *The Economic Journal*, 116(510): C63–C76.

Hanushek, Eric A, and Margaret E Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, 24(2): 297–327.

Harris, Milton, and Bengt Holmström. 1982. "A Theory of Wage Dynamics." *Review of Economic Studies*, 49(3): 315–333.

Hastings, Justine, Christopher A Neilson, and Seth D Zimmerman. 2015. "The Effects of Earnings Disclosure on College Enrollment Decisions." National Bureau of Economic Research Working Paper No. 21300.

Hastings, Justine, C. Neilson, and S. Zimmerman. 2013*a*. "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile." National Bureau of Economic Research Working Paper No. 19241 Mimeo.

Hastings, Justine S, and Jeffrey M Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *The Quarterly Journal of Economics*, 123(4): 1373–1414.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman. 2013*b*. "Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile." National Bureau of Economic Research Working Paper 19241.

Heckman, James J, Lance Lochner, and Christopher Taber. 1998. "General-Equilibrium Treatment Effects: A Study of Tuition Policy." *American Economic Review*, 88(2): 381–86.

Heisz, Andrew, and Phil Oreopoulos. 2002. "The importance of signaling in job placement and promotion." Mimeo, University of California at Berkeley.

He, Yingha. 2014. "Gaming the Boston School Choice Mechanism in Beijing." Mimeo, Toulouse School of Economics.

Hinrichs, Peter. 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, Peter. 2014. "Affirmative Action Bans and College Graduation Rates." *Economics of Education Review*, 42: 43–52.

Hoekstra, Mark. 2009*a*. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-based Approach." *The Review of Economics and Statistics*, 91(4): 717–724.

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

Hoxby, Caroline. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives*, 23(4): 95–118.

Hoxby, Caroline, and Christopher Avery. 2012. "The missing 'one-offs': The hidden supply of high-achieving, low income students." National Bureau of Economic Research Working Paper 18586.

Hoxby, Caroline, and Christopher Avery. 2013*a*. "The missing 'one-offs': The hidden supply of high-achieving, low income students." *Brookings Papers on Economic Activity*, Spring: 1–65.

Hoxby, Caroline, and Sarah Turner. 2013*a*. "Expanding college opportunities for high-achieving, low income students." Working Paper 12-014.

Hoxby, Caroline, and Sarah Turner. 2013*b*. "Expanding College Opportunities for High-achieving, Low Income Students." Stanford Institute for Economic Policy Research Discussion Paper.

Hoxby, Caroline M. 1997. "How the Changing Market Structure of U.S. Higher Education Explains College Tuition." National Bureau of Economic Research Working Paper No. 6323.

Hoxby, Caroline M. 2003. "School Choice and School Productivity. Could School Choice Be a Tide that Lifts All Boats?" In *The Economics of School Choice.* , ed. Caroline M Hoxby, 287–342. University of Chicago Press.

Hoxby, Caroline M, and Christopher Avery. 2013*b*. "Missing One-Offs: The Hidden Supply of High-Achieving, Low-Income Students." Brookings Papers on Economic Activity.

Hubbard, Thomas N. 2002. "How do consumers motivate experts? Reputational incentives in an auto repair market." *Journal of Law & Economics*, 45(2): 437–468.

Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics*, 86(1): 4–29.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *The Quarterly Journal of Economics*, 125(2): 515–548.

Jin, Ginger Zhe, and Phillip Leslie. 2003. "The effect of information on product quality." *Quarterly Journal of Economics*, 118(2): 409–451.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy*, 87: 972–90.

Kahn, L. B., and F. Lange. 2014. "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures." *Review of Economic Studies*, 81(4): 1575–1613.

Kane, Thomas J, and Douglas O Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *The Journal of Economic Perspectives*, 16(4): 91–114.

Kapor, Adam. 2015. "Distributional Effects of Race-Blind Affirmative Action." Working Paper.

Katz, Lawrence F, and Kevin M Murphy. 1992*a*. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, 107(1): 35–78.

Katz, Lawrence F, and Kevin M Murphy. 1992*b*. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, 107(1): 35–78.

Kaufmann, Katja M, Matthias Messner, and Alex Solis. 2013. "Returns to elite higher education in the marriage market: Evidence from Chile." Bocconi University Mimeo.

Ketel, Nadine, Edwin Leuven, Hessel Oosterbeek, and Bas van der Klaauw. 2012. "The Returns to Medical School in a Regulated Labor Market: Evidence from Admission Lotteries." Working Paper.

Kirkebøen, Lars, Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *The Quarterly Journal of Economics*, 131(3): 1057–1111.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica*, 56(2): 259–294.

Lange, Fabian. 2007. "The Speed of Employer Learning." *Journal of Labor Economics*, 25(1-35): 497–532.

Lemieux, T. 2006. "The Mincer equation thirty years after schooling, experience, and earnings." In *Jacob Mincer, A pioneer of modern labor economics.* , ed. S. Grossbard, 127–145. Springer Verlag.

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review*, 105(11): 3471–3488.

MacLeod, W Bentley, Evan Riehl, Juan E Saavedra, and Miguel Urquiola. 2015. "The Big Sort: College Reputation and Labor Market Outcomes." *American Economic Journal: Applied Economics* (forthcoming).

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Econ*, 60(3): 531–542.

Manski, Charles F. 2013. "Identification of Treatment Response with Social Interactions." *The Econometrics Journal*, 16(1): S1–S23.

Martorell, Paco, and Isaac McFarlin Jr. 2011. "Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes." *The Review of Economics and Statistics*, 93(2): 436–454.

Melnik, Mikhail I., and James Alm. 2002. "Does a Seller's Ecommerce Reputation

Matter? Evidence from eBay Auctions." *Journal of Industrial Economics*, 50(3): 337–49.

Milkovich, George T., Jeffrey M. Newman, and Barry Gerhart. 2011. *Compensation.* . Tenth Edition ed., Chicago, IL:Irwin.

Mincer, Jacob. 1974. *Schooling, experience and earnings.* New York:Columbia University Press.

Murphy, Richard, and Felix Weinhardt. 2014. "Top of the Class: The Importance of Ordinal Rank." CESifo Working Paper Series.

Nelson, Phillip. 1970*a*. "Information and consumer behavior." *Journal of Political Economy*, 78(2): 311–329.

Nelson, Phillip. 1970*b*. "Information and Consumer Behavior." *Journal of Political Economy*, 78(2): 311–29.

Nguyen, Trang. 2010. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." Unpublished manuscript.

Oreopoulos, Philip, and Ryan Dunn. 2013. "Information and College Access: Evidence from a Randomized Field Experiment." *The Scandinavian Journal of Economics*, 115(1): 3–26.

Pallais, Amanda. 2015. "Small Differences That Matter: Mistakes in Applying to College." *Journal of Labor Economics*, 33(2): 493–520.

Perez-Arce, Francisco. 2015. "Is a Dream Deferred a Dream Denied? College Enrollment and Time-Varying Opportunity Costs." *Journal of Labor Economics*, 33(1): 33–61.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review*, 103(4): 1289–1324.

Ramey, Garey, and Valerie A Ramey. 2010. "The Rug Rat Race." *Brookings Papers on Economic Activity*, 41(1 (Spring)): 129–199.

Rao, Gautam. 2013. "Familiarity does not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools." Job Market Paper.

Riehl, Evan. 2015. "Competition in College Enrollment: The Effects of an Academic Calendar Shift in Colombia." Columbia Center for Development Economics and Policy (CDEP) and Center on Global Economic Governance (CGEG) Working Paper No. 20.

Rivkin, Steven G, Eric A Hanushek, and John F Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417–458.

Rosen, Sherwin. 1981. "The Economics of Superstars." *The American Economic Review*, 71(5): pp.845–858.

Roth, A. E., and M. Sotomayor. 1989. "The College Admissions Problem Revisited." *Econometrica*, 57(3): 559–570.

Rothschild, Michael, and Lawrence J White. 1995. "The analytics of the pricing of higher education and other services in which the customers are inputs." *Journal of Political Economy*, 573–586.

Rothstein, Jesse, and Albert Yoon. 2008. "Mismatch in Law School." National Bureau of Economic Research Working Paper No. 14275.

Rothstein, Jesse M. 2004. "College Performance Predictions and the SAT." *Journal of Econometrics*, 121(1): 297–317.

Saavedra, Juan Esteban. 2009. "The Returns to College Quality: A Regression Discontinuity Analysis." Harvard University.

Sahin, A., J. Song, G. Topa, and G. L. Violante. 2014. "Mismatch unemployment." *American Economic Review*, 104(11): 3529–3564.

Schütz, Gabriela, Heinrich W Ursprung, and Ludger Woessmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos*, 61(2): 279–308.

Scott-Clayton, Judith E. 2011. "The Shapeless River: Does a Lack of Structure Inhibit Students' Progress at Community Colleges?" Community College Research Center Working Paper.

Siow, Aloysius. 2015. "Testing Becker's Theory of Positive Assortative Matching." *Journal of Labor Economics*, 33(2): 409–441.

Spence, Michael. 1973. "Job market signaling." *The Quarterly Journal of Economics*, 3: 355–374.

Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy*, 69(3): 213–225.

Stinebrickner, Ralph, and Todd 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.

Tincani, Michela M. 2014. "Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence." Job Market Paper.

Topel, R. H., and M. P. Ward. 1992. "Job mobility and the careers of young men." *The Quarterly Journal of Economics*, 107(2): 439–479.

Topkis, Donald M. 1998. *Supermodularity and Complementarity.* Princeton University Press.

Urquiola, Miguel. 2016. "Competition Among Schools: Traditional Public and Private Schools."

Urzua, Sergio, J. Rodriguez, and L. Reyes. 2015. "Heterogenous Economic Returns to Postsecondary Degrees: Evidence from Chile." University of Maryland Mimeo.

Venezia, Andrea, Michael Kirst, and Anthony Antonio. 2003. "Betraying the College Dream: How Disconnected K-12 and Postsecondary Education Systems Undermine Student Aspirations." Palo Alto, CA: Stanford University.

Wiswall, Matthew, and Basit Zafar. 2015. "Determinants of College Major Choice: Identification using an Information Experiment." *The Review of Economic Studies*, 82(2): 791–824.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review*, 93(2): 133–138.

Zafar, Basit. 2011. "How do College Students form Expectations?" *Journal of Labor Economics*, 29(2): 301–348.

Zimmerman, Seth. 2013. "Making top managers: The role of elite universities and elite peers." Yale University Mimeo.

Zimmerman, Seth. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics*, 32(4): 711–754.