THE BIG SORT: COLLEGE REPUTATION AND LABOR MARKET OUTCOMES

W. BENTLEY MACLEOD*
EVAN RIEHL+
JUAN E. SAAVEDRA#
MIGUEL URQUIOLA*

MAY 25, 2015

Abstract. Spence (1973) noted that individuals' choice of educational *quantity*—measured by years of schooling—may stem partially from a desire to signal their ability to the labor market. This paper asks if individuals' choice of educational *quality*—measured by college reputation—may likewise signal their ability. We use data on the admission scores of all Colombian college graduates to define a measure of reputation that gives clear predictions in a signaling framework. We find that college reputation, unlike years of schooling, is correlated with graduates' earnings *growth*. We also show that Colombia's staggered rollout of a new signal of skill—a college exit exam—reduced the earnings return to reputation and increased the return to individual admission scores. These results are consistent with the hypothesis that a college's reputation provides information about the ability of its student body and about its value added, broadly understood.

For useful comments we thank Joseph Altonji, Costas Meghir, Michael Mueller-Smith, Phil Oreopoulos, and Kiki Pop-Eleches. For invaluable help with the data we are 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 ours.
*Columbia University and NBER, +Columbia University, #University of Southern California.

1. Introduction

For many individuals, whether to enroll in college is a major decision. For others who are clearly college-bound, the question of which college to attend absorbs substantial energy.[1] The difference between these two choices—"whether college" and "which college"—is analogous to the distinction between *quantity* and *quality* in industrial organization. Spence (1973) introduced the idea that individuals may use the quantity (years) of education to signal their ability; if it is less costly for more able individuals to consume education, they will acquire more schooling and make higher earnings as a result.

This paper asks if individuals' choice of educational quality may likewise signal ability. We explore this possibility in the context of college identity using administrative data on all graduates in the country of Colombia. We introduce a simple measure of college quality and analyze how it correlates with individuals' starting earnings and earnings growth. Finally, we ask how the relationship between college quality and earnings changed with the staggered introduction of a new source of information on worker skill—a national college exit exam.

A standard observation in industrial organization is that the quality of a good is more difficult to ascertain than its quantity.[2] This is even more so for experience goods like education. The quality of a college, for instance, ultimately depends on the experience one has there, as well as on career outcomes realized many years later. Since students cannot try out colleges before making a decision, they must choose based on expectations that are shaped by colleges' *reputations*.

A college's reputation is a complex, multidimensional signal that is less readily quantifiable than Spence's years of schooling. This complicates analysis of its signaling function because any measure of reputation chosen by researchers is likely to be imperfectly observed by employers. In this paper we adopt a simple definition of a college's reputation: the average admission test score of its graduates. This is a reasonable proxy for reputation in countries like Colombia, where selective colleges use test scores to determine admission. If students prefer more "reputable" colleges there will be a positive mechanical relationship between colleges' admission scores and other desirable attributes.

We extend the theoretical literature on information and wage formation to show that our measure of reputation provides clear testable predictions related to signaling. The framework covers two distinct empirical analyses. The first relates to Farber and Gibbons' (1996) and Altonji and Pierret's (2001) finding that as workers gain experience, observable characteristics like years of schooling become less correlated with wages in regressions that include unobserved measures of ability. This suggests that schooling transmits information on ability,

[1] In many countries significant industries surround college admissions, with elements ranging from standardized testing preparation to other types of advising (Dang and Rogers, 2008; Ramey and Ramey, 2010).
[2] See Nelson (1970) for a seminal discussion of information and consumer choice.

while human capital growth and other factors correlated with schooling have a deterministic effect on wages. In other words, schooling influences initial wages but is not related to wage growth.[3]

The prediction for schooling does not immediately extend to college reputation if it is imperfectly observed by employers. However, defining reputation as college-mean admission score allows us to circumvent this issue. This is because the reputation of an individual's college of graduation is mechanically a noisy version of her own admission score. Thus if college reputation only signals ability as measured by admission scores, the coefficient on reputation in a wage equation should similarly decline over time in a regression that includes individual scores.

We test this using data that link individuals' standardized admission exam scores, college identity, and formal labor market earnings. We find robust evidence that the above prediction does not hold. Even controlling for individual admission scores, college reputation is positively correlated not only with graduates' initial earnings but also with their earnings *growth*—a starkly different pattern than that observed for years of schooling.[4]

The increasing correlation between reputation and earnings could arise if colleges offer different amounts of value added, human capital externalities, or network effects. It could also reflect employer learning about other attributes, like family income, that influence students' college choices and are not perfectly correlated with admission scores. In either case, our findings suggest that the sorting that takes place by educational quality differs from that which takes place by quantity. In short, our results are inconsistent with college identity performing *only* a signaling function.

Our second empirical analysis shows that reputation does play a signaling role. For this we exploit the Colombian government's introduction of national college exit exams that provide the labor market with an individual-specific measure of skill. Our theoretical framework illustrates that the exit exams should decrease the correlation of earnings with reputation, and increase their correlation with ability as measured by admission scores. The intuition is twofold. First, if employers observe an independent signal of skill, they may rely less on college reputation to infer it. Second, the exit exam is correlated with admission scores and thus allows employers to more quickly observe a trait that gradually becomes more correlated with earnings.

The exit exams were rolled out in a staggered fashion across 55 fields like accounting, dentistry, economics, and law. This led to different levels of exam-taking across college programs.[5] We exploit this variation to implement an approach analogous to Card and

[3] Equivalently, schooling changes only the intercept of earnings-experience profiles (Lemieux, 2006).
[4] Equivalently, reputation is correlated with both the intercept and the slope of earnings-experience profiles.
[5] Throughout the paper we refer to college majors as "programs" and exit exam subjects as "fields."

Krueger (1992), who analyze how time-varying state policies (e.g., class size levels) can affect a slope—the relation between years of schooling and wages. In our case the question is how time-varying college program characteristics (e.g., the existence of an exit exam in a related field) can affect two slopes—the earnings return to reputation and the earnings return to admission scores. We find that the exit exams reduced the return to reputation and increased the return to admission scores, thus suggesting that college reputation transmits information on ability. In addition, we present suggestive evidence that the exit exams improved overall match quality as measured by average earnings, and that they prompted behavioral responses in the form of delayed graduation and preference for colleges and programs with better exit exam performance.

We note two potential threats to identification in this analysis. First, due to data restrictions we do not observe pre-treatment cohorts at very early experience levels. This constrains how flexibly we can explore the experience profile of earnings and how it interacts with the availability of the exit exam. Second, our central identification assumption is one of parallel trends; we assume that the returns to reputation and to admission scores would have evolved similarly across programs in the absence of the exit exam. This is not innocuous because the programs that received exams earlier and later are not otherwise identical. A series of robustness checks nonetheless confirm our main findings.

The bottom line is that educational quality matters in the "big sort" that occurs as students transition into college and out to the labor market. The signaling role of school reputation provides one explanation for the substantial resources devoted to college admissions in many countries. In addition, the correlation between reputation and earnings growth may lead college applicants to perceive that school identity matters for their careers—that going to a better school can put them on a different earnings trajectory (Topel and Ward, 1992).

This paper is related to various literatures. Most directly, it addresses the possibility that college reputation can perform a signaling function, as modeled by MacLeod and Urquiola (2013). This complements research on the earnings impact of attending selective colleges (Dale and Krueger, 2002, 2014; Hoekstra, 2009; Saavedra, 2009; Hastings et al., 2013; Urzua et al., 2015). Our contribution is to explicitly measure reputation and relate it to earnings in a whole market. Our results suggest that information-related mechanisms may account for some of the effects in this literature. On the other hand, our findings do not foreclose other mechanisms, such as human capital externalities (Epple et al., 2006) or better networks (Kaufmann et al., 2013; Zimmerman, 2013a). Further, we illustrate that earnings effects may be setting-specific and depend on the experience levels at which they are measured.

Our paper also informs research on how sorting in educational systems relates to the labor market. While we document extensive stratification by ability in Colombia, this sorting is far from complete. Costs still significantly influence college choice, and as a result college markets in Colombia today tend to resemble the regional autarkies observed in the U.S. in the 1970s (Hoxby, 1997, 2009). This is consistent with our finding that college identity does not fully reveal admission scores in Colombia, whereas it seems to fully convey Armed Forces Qualification Test scores in the U.S. (Arcidiacono et al., 2010).

Our work also relates to research on how the configuration of educational systems affects human capital investment. For example, Coate and Loury (1993), Bishop (2004), and MacLeod and Urquiola (2013) suggest that the rules governing the allocation of educational or job opportunities may affect the incentives that students have to invest effort into human capital accumulation. Our results are generally consistent with such mechanisms, and additionally suggest that exit exams may enhance job matching (Biglaiser, 1993).

The remainder of the paper proceeds as follows. Section 2 extends the benchmark model of information and wage formation to incorporate college reputation. Section 3 documents that college reputation is correlated with earnings growth. Section 4 describes the introduction of the exit exams and their effect on the correlation of earnings with reputation and admission scores. Section 5 concludes.

2. College reputation, signaling, and wages

This section extends the literature on information and wage formation (Jovanovic, 1979; Farber and Gibbons, 1996; Altonji and Pierret, 2001) to incorporate college reputation. We present the basic theory and state two empirical predictions that we test in Sections 3 and 4. A full derivation of the model and propositions appears in Appendix A.

2.1. 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 that we can define two measures of α_i in our data. First, we observe each student's score on a college admission exam, denoted by τ_i, and assume it provides a noisy measure of ability:

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

Our second measure of ability is the quality of a student's college. As education is an experience good, students' decisions of where to enroll are shaped by expected outcomes based on colleges' *reputations*. A college's reputation may incorporate many aspects of its quality, including peer composition, faculty research output, and financial resources. This makes it challenging to define a measure of reputation that is conducive to empirical analysis.

In this paper 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 definition has two analytical advantages. First, in settings where selective colleges use test scores to determine admission, our reputation measure will have a positive mechanical relationship with other dimensions of quality that lead students to prefer certain colleges. Second, as we discuss below, defining reputation as college-mean admission score delivers clear predictions related to signaling 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 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:

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

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:

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

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 education and revealed over time to the employer.

We suppose that employers' information set, I_{it}, includes our measure of college reputation, R_{s_i}. This is a strong assumption given its finely-grained nature and the ambiguities in defining college quality. Employers likely observe college *identity* through students' CVs, but they may not have access to our measure of reputation defined by college-mean admission scores. We make this assumption to derive benchmark predictions consistent with the literature on observable characteristics like years of schooling. Below we discuss the empirical implications if employers do not perfectly observe R_{s_i}, and how our definition of reputation helps to address this issue.

Our measure of reputation, R_s, captures the pre-college ability of individuals at college s, i.e., the quality of its "student inputs." In setting wages employers are rather interested in 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 $\mathscr{R}_s$.[6]

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 is the reputation premium, i.e., the return to reputation beyond that captured by admission scores. If this reputation premium is positive (if $v_1 > 0$) then from a student's point of view a college with a better reputation provides higher value added, broadly understood.

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

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

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:

$$y_i = \alpha_i + v_0 + v_1 R_{s_i} + \epsilon_i,$$

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 Armed Forces Qualification Test scores

[6] 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.

are unobserved, and with anecdotal evidence that in our setting graduates' CVs rarely feature their college admission exam score.

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

$$y_{it} = \alpha_i + v_0 + v_1 R_{s_i} + \epsilon_{it},$$

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 $\rho^{\bar{y}}$.[7]

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

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

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 the new information from worker output.

2.3. **Predictions for earnings growth.** Equation (1) 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. In the empirical sections 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 of the model for the relationship between the signals of individual ability and wages net of human capital growth.

We define the unconditional *return to reputation* at time t, r_t^u, and the unconditional *return to ability*, a_t^u, as the coefficients from the regressions:

$$\hat{w}_{it} = r_t^u R_{s_i} + e_{it}^R \tag{2}$$

$$\hat{w}_{it} = a_t^u \tau_i + e_{it}^{\tau}, \tag{3}$$

where the e_{it} variables are residuals. The return to reputation, r_t^u, is the coefficient on our measure of college reputation, R_s, and the return to ability, a_t^u, is the coefficient on the admission exam score, τ_i.

If colleges are not perfectly selective, we can identify *conditional* returns to college reputation, r_t, and to ability, a_t, from the regression:

$$\hat{w}_{it} = r_t R_{s_i} + a_t \tau_i + e_{it}. \tag{4}$$

[7] The assumption that the precision of $\bar{y}_{it}$ is time stationary also follows Farber and Gibbons (1996).

In this specification, r_t is the return to reputation conditional on admission scores—the earnings impact of a change in reputation for students with similar academic ability—while a_t measures the return to ability for students from schools with similar reputations.

While the returns to reputation and to ability are not causal, changes in these parameters are informative as to the signaling role of reputation. In Section 3, we describe how these returns change with experience, t, thereby comparing college reputation to other potential signals of ability studied in the literature. In Appendix A.3 we show formally that the evolution of the regression coefficients from (2)-(4) satisfy the following proposition:

Proposition 1. *If wages are set equal to expected skill given the available information (equation* (1)*), 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.*

Proposition 1 shows that the basic information model provides a rich set of testable implications. Parts (1)-(2) mirror the Farber and Gibbons (1996) predictions that observable characteristics are fully incorporated in initial wages, while employers gradually learn about unobservable traits through signals of worker output. In our model, reputation, R_s, has a constant effect over time because we assume that it is observed by the labor market, 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}. If reputation is imperfectly observed, its unconditional return should similarly grow with experience, a possibility we discuss in the empirical analysis below.

Parts (3)-(4) predict a declining conditional return to reputation, and an increasing conditional return to ability. These match the Altonji and Pierret (2001) predictions for observable and unobservable characteristics, but our definition of R_s makes this prediction an even stronger test of the role of reputation in signaling. Since we define reputation as the mean admission score at a college, admission scores are a sufficient statistic for ability, α_i, in regression (4). This means that part (3) of Proposition 1 holds even if employers imperfectly observe our measure of reputation, or if there are interactions between α_i and human capital growth; all of these effects are captured in the admission score coefficients in the conditional returns regression (4). 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 1 allows us to test whether the return to reputation arises solely because college identity signals ability as measured by admission scores. This is akin to the classic Spence hypothesis in the context of educational quality rather than educational quantity. 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 3.

2.4. **Predictions for the introduction of a college exit exam.** In Section 4 we ask how the conditional returns to reputation and ability from equation (4) 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 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.

The increase in the precision of y_i has the effect of reducing the weight on reputation in wage setting, $\pi_t^{\mathscr{R}}$, for every t. Let $\delta_i = 1$ if and only if a student is exposed to the possibility of writing the exit exam. We can rewrite regression (4) as follows:

$$
\begin{aligned}
\hat{w}_{it} &= (1-\delta_i)\left(r_t R_{s_i} + a_t \tau_i\right) + \delta_i \left(r_t^{exit} R_{s_i} + a_t^{exit}\tau_i\right) + e_{it}^{exit} \\
&= \left(r_t R_{s_i} + a_t \tau_i\right) + \delta_i \left(\beta_t^r R_{s_i} + \beta_t^a \tau_i\right) + e_{it}^{exit},
\end{aligned}
\tag{5}
$$

where $\beta_t^r = r_t^{exit} - r_t$ and $\beta_t^a = a_t^{exit} - a_t$. In Appendix Section A.4 we show that $\beta_t^r < 0$ and $\beta_t^a > 0$.[8] Thus we have:

Proposition 2. *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 2 yields a different test of the role of college reputation in transmitting information on ability. If employers do not use reputation in setting 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).

This test is particularly strong if there is exogenous variation in access to the exit exam, captured by the indicator δ_i. In Section 4, we use the staggered introduction of field-specific graduation exams to test Proposition 2.

[8] The appendix also shows that the exit exams should have no effect on the *unconditional* return to reputation, r_t^u, and a positive but smaller effect on the unconditional return to ability, a_t^u.

3. College reputation and earnings growth

This section shows how college reputation correlates with initial earnings and with earnings growth in Colombia. We first describe the institutional background and our data sources. We then calculate a measure of college reputation from individual admission scores and use both measures to test Proposition 1 from Section 2.

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 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.[9]

To gain admission to college, Colombian students are required to take a standardized admission exam, the Icfes, which is administered by a government agency.[10] 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.[11] The Icfes also plays a larger role in admissions in Colombia than the SAT does in the U.S. In addition to being mandatory for application to any college, many schools extend admission offers based solely on students' Icfes performance; others consider additional factors like high school grades while heavily weighting the Icfes, 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 (see footnote 10) provided Icfes scores for all high school seniors who took the exam between 1998 and 2012. This agency also provided college exit exam scores for all students who took the exam in 2004–2011 (see Section 4).
(2) The Ministry of Education provided enrollment and graduation records for students entering college between 1998 and 2012. These include enrollment date, graduation

[9] The Colombian Ministry of Education classifies higher education institutions into five types: universities, university institutes, technology schools, technology institutes, and technical/professional institutes. We define the first two as colleges; these enroll over 90 percent of post-secondary students in our records. The Ministry also categorizes programs based on their normative duration: university-level (four to five years), technological (three years), or technical professional (two years). We focus on students from university-level programs, which comprise about 80 percent of all programs.

[10] Icfes stands for Institute for the Promotion of Higher Education, the former acronym for the agency that administers the exam. The Colombian Institute for Educational Evaluation, as it is now called, was created in 1968 and is a State agency under the authority of the national Ministry of Education. The Icfes exam is now known as Saber 11°, reflecting the fact that students usually take it in the 11^{th} grade. We use the name Icfes to match the designation during the period covered by our data.

[11] Angrist et al. (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.

or dropout date, program of study, college, and each student's 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 data are derived from information on contributions to pension and health insurance funds. From these we calculate average daily earnings for each year by dividing base monthly earnings for pension contributions by the number of formal employment days in each month and averaging across months.[12] This agency also provided four-digit economic activity codes for the first job in which a worker appears in their records.

3.2. **Ability sorting and college reputation.** Theoretical research suggests that the Colombian college system has characteristics that facilitate sorting by ability: relatively free entry by private colleges, choice on the part of students, and a recognized measure of ability.[13] At the same time, the costs of college might counteract this tendency. Tuition is significant for many households given that educational credit markets and financial aid mechanisms are less developed than in the U.S., and the costs of moving away for college may still be substantial for some families. Consistent with this, college students often live with their parents, and dormitories are still rare.[14] Hoxby (1997) shows that such cost barriers can reduce ability sorting, leading to variation in ability within colleges.

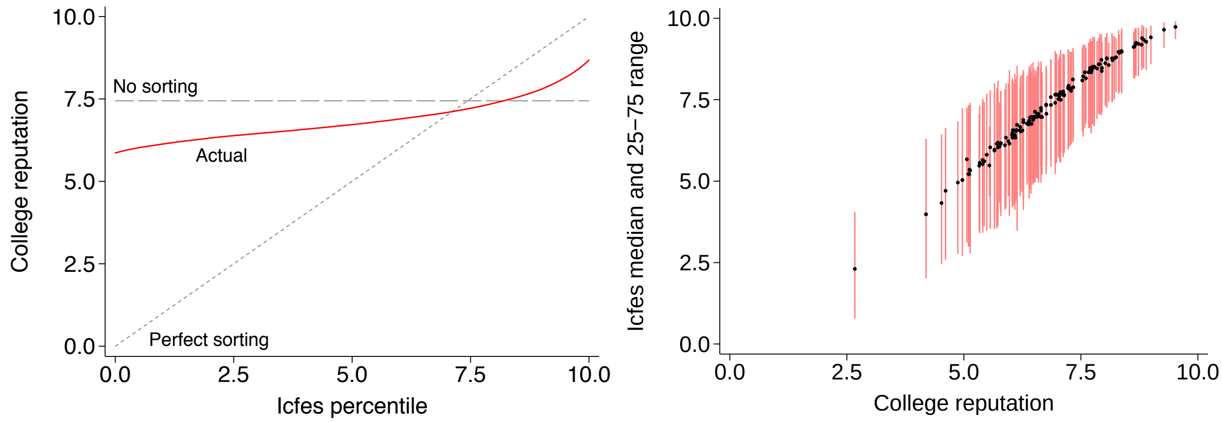
We study these sorting patterns using two measures of ability from the theoretical framework in Section 2. The first is student i's score on the Icfes admission exam, which we denote by τ_i. Throughout the paper, we express Icfes scores as percentiles relative to all 11^{th} grade test takers in the same exam year. Second, we use the reputation of a college s, denoted by R_s, which as above is the mean Icfes score of its graduates. We compute reputation measures for the 136 Colombian colleges that have at least ten graduates per Icfes cohort. We divide both Icfes percentiles and reputation by ten so that both measures have potential ranges of 0–10. A one unit increase in Icfes is therefore equivalent to ten percentile points in the full distribution of exam takers, while a one unit increase in reputation indicates that a college's average graduate scored ten percentile points higher on the Icfes.

Figure 1 describes ability sorting in Colombia using these two measures. Panel A presents the college reputation (y-axis) experienced by college graduates as a function of their Icfes scores (x-axis). The dotted lines describe two polar cases of across-college sorting by ability.

[12] In this paper we use earnings related to pension contributions, but our main results are nearly identical when we use earnings based on health contributions (see Column (B) in Appendix Table B7).

[13] Models like Epple and Romano (1998), Epple et al. (2003), and MacLeod and Urquiola (2013) suggest that relatively unfettered educational markets will tend towards stratification.

[14] Saavedra (2012) reports that over 70 percent of students attend college in their state of their birth.



Panel A. Across colleges

Panel B. Within colleges

FIGURE 1. Distribution of ability across and within colleges

Notes: Panel A 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 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. The no sorting line is defined by the mean Icfes percentile among all graduates (74.4 percent). The perfect sorting line assumes each student graduates from a college whose reputation equals her Icfes score. The actual line is the predicted value from a local linear regression of college reputation on Icfes percentiles with Stata's default bandwidth.

Panel B displays the 136 colleges included in Panel A by their reputation. Black dots are the median Icfes percentiles among graduates from each school, and vertical lines are the 25^{th}–75^{th} Icfes percentile ranges.

First, the horizontal line depicts a case in which all students graduate from colleges with the same reputation; this would result, for instance, if students were randomly allocated to colleges. Since many students who take the Icfes never enroll in or graduate from college, this line is at the 74^{th} percentile—the national mean Icfes percentile of graduates. Second, the 45 degree line depicts a perfect sorting benchmark; this would result, for instance, if there were a large number of colleges and each only admitted students with a given Icfes score. The solid curve indicates the observed distribution, smoothed using a local linear regression. It shows significant but incomplete sorting by ability. The mean college reputation increases by more than 25 percentile points across the distribution of Icfes scores, but this relationship is much flatter than the perfect sorting benchmark.

This incomplete sorting results in significant variation in Icfes scores within each school. We illustrate this in Panel B. The horizontal axis depicts the reputation of the 136 colleges from Panel A. 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} Icfes percentile ranges. Panel B shows a large mass of colleges near the middle of the distribution and fewer near the high and low 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.

The substantial variation in ability both across and within colleges allows us to estimate distinct earnings returns to college reputation and Icfes, a task to which we now turn.

3.3. **Empirical results on earnings growth.** Section 2 described how log wages relate to characteristics in our data under a basic information framework. This section tests Proposition 1 from Section 2, which predicts how coefficients on college reputation and Icfes scores change with worker experience.

3.3.1. *Sample.* 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. We focus on students who graduated from college in 2008 or 2009. 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. In addition, we consider only students who entered the labor market immediately upon graduation and remained in it during four consecutive years; we exclude students who attended graduate school or were not formally employed in any year after graduation.[15] The results are thus not attributable to movements into and out of employment; they originate in earnings changes within the formal labor market. The students in our sample represent about one quarter of all graduates and are slightly higher ability on average. Appendix B.1 provides further details on the construction and characteristics of the sample.

3.3.2. *Empirical specifications.* Below we estimate regressions that test Proposition 1 (Section 2). Our basic specification is:

$$w_{it} = d_{c_i t} + r_0 R_{s_i} + r \left(R_{s_i} \times t\right) + a_0 \tau_i + a \left(\tau_i \times t\right) + e_{it}. \tag{6}$$

The dependent variable, w_{it}, is log daily earnings for student i measured at potential experience t, which we define as employment year minus graduation year.[16] All regressions include dummies for cells defined by graduation cohort c_i and experience t, denoted by $d_{c_i t}$. We report only coefficients on reputation, R_{s_i}, Icfes, τ_i, and their interactions with experience. The r_0 coefficient is the *return to reputation* in the year of graduation, while r represents the average change in the return to reputation from an additional year of potential labor market

[15] Specifically, our sample includes students who do not appear in our 2007–2011 graduate education records, and who have at least one monthly earnings observation in each of the first four years after graduation.

[16] 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 suggests that hours are relatively constant early in college graduates' careers, and, if anything, decline with ability as measured by years of schooling. This suggests that the results below are not driven by the difference between hourly wages and daily earnings.

experience.[17] Similarly, a_0 is the period-zero *return to ability*, and a represents the average yearly change in this return. As stated the coefficients on the experience interactions, r and a, 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 specification (6), our goal is not to identify the causal effect of reputation or admission scores. The return to reputation, r_0, is analogous to the college wage premium—the average difference in earnings between college and high school graduates. The college premium measures the descriptive average return in the "whether college" dimension, while the return to reputation does the same in the "which college" dimension.[18] This measure incorporates numerous individual and school characteristics that vary across colleges. Similarly, a_0 is a population parameter capturing the average change in earnings from an increase in ability as measured by Icfes. Our interest is in how these population returns change with worker experience—the r and a coefficients—and whether these changes match the predictions from the signaling model of Section 2.

Table 1 estimates equation (6) including the reputation and Icfes terms both separately and jointly. This corresponds to regressions (2), (3), and (4) from Section 2, which yield the unconditional return to reputation, the unconditional return to ability, and the conditional returns to both characteristics. We discuss results from each of the three regressions separately in the subsections below.

3.3.3. *Unconditional return to reputation.* Column (A) of Table 1 estimates (6) including reputation terms but not Icfes terms, such that the estimates represent the unconditional return to reputation, r^u (equation (2), Section 2). 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 ($r_0^u \approx 0.10$). 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."

Proposition 1 predicts that the unconditional return to reputation should not change with experience, which implies a zero coefficient on the interaction of reputation and experience, t. This arises because initial wages fully incorporate information observable to employers, which we assume includes college reputation. Reputation therefore cannot predict innovations in wages, which depend on signals related to the worker's output. This is identical to wages being a martingale in Farber and Gibbons (1996).

[17] Formally, we parametrize the experience-specific r_t coefficients in equation (4) as $r_t = r_0 + r \times t$.
[18] Many papers analyze the college wage premium and its properties. For example, Katz and Murphy (1992) and Card and Lemieux (2001) consider the evolution of the college wage premium in the 1980s.

Table 1. Returns to reputation and ability and experience interactions
Dependent variable: log average daily earnings

	(A)	(B)	(C)
Reputation	0.101***		0.079***
	(0.017)		(0.017)
Reputation × t	0.017***		0.012***
	(0.003)		(0.003)
Icfes		0.045***	0.024***
		(0.006)	(0.002)
Icfes × t		0.009***	0.006***
		(0.001)	(0.001)
N	83,492	83,492	83,492
R^2	0.179	0.163	0.190
# colleges	130	130	130

Notes: The dependent variable is log average daily earnings (see Section 3.1). The sample includes students in column (D) of Appendix Table B1 and earnings in the four years after graduation. Columns (A)-(C) estimate equation (6) including, respectively: only reputation terms, only Icfes terms, and all terms. In addition to the reported variables, all regressions include dummies for cohort-experience cells. Parentheses contain standard errors clustered at the college level.

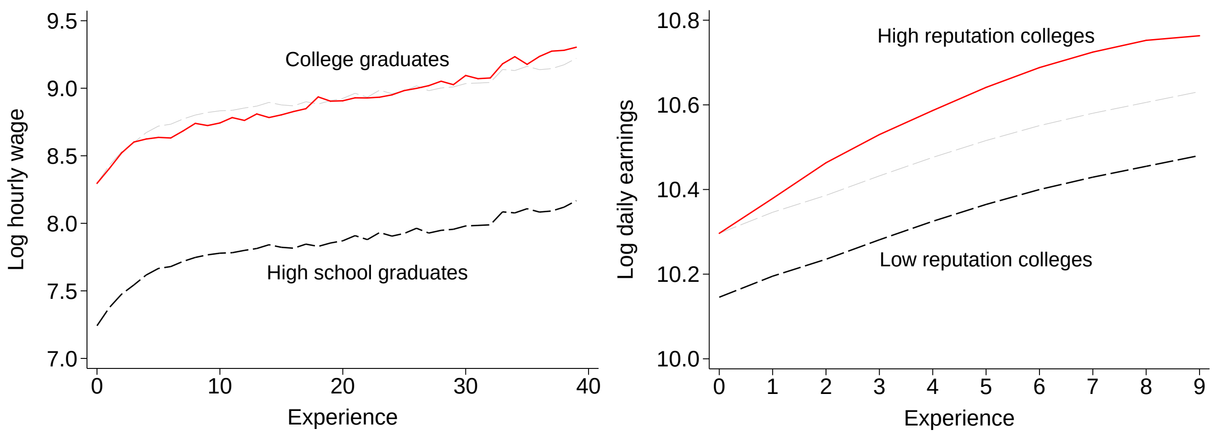
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.[19]

The contrast between the reputation and years of schooling results can also be depicted using earnings-experience profiles. Mincer (1974) noted that wage profiles of workers with different schooling levels are approximately parallel throughout the earnings lifecycle. Panel A of Figure 2 replicates this finding using 2008–2012 household survey data from Colombia.[20] We plot the mean log hourly real wage among workers with two schooling levels—completed high school and completed college—so that the gap between the two profiles is the college

[19] Altonji and Pierret (2001) find that another potential signal of ability—race—does have an increasing relationship with wages. They attribute this to legal restrictions on the market's use of race in setting initial wages. In our setting there are no legal barriers to statistical discrimination on the basis of college reputation.

[20] In Figure 2, 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) 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).



Panel A. Years of schooling

Panel B. College reputation

Figure 2. 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 $\text{age} - 16 - 6$. The dashed light grey line is parallel to the low reputation profile starting from the high reputation intercept.

wage premium. This gap remains roughly constant across forty years of potential experience, consistent with the standard Mincerian result in the U.S. (Lemieux, 2006).[21]

Panel B uses our administrative data to define 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, which allows us to observe earnings up to nine years after graduation.[22] 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

[21] Appendix Table B2 reports the results from a regression of log wages on a linear term in years of schooling, experience, and their interaction using the Colombian survey data. Consistent with Panel A of Figure 2, the coefficient on the interaction of schooling and experience is approximately zero.

[22] In other words, the sample for Panel B differs from that in Table 1, and each worker contributes five years of earnings observations at most. To be consistent with Panel A, we also use experience defined by age and years of schooling. We assume all college graduates have 16 years of schooling, and thus in Panel B experience is $\text{age} - 16 - 6$. With this definition we can actually observe levels of experience above nine years, but we omit these levels because they appear only for workers who took especially long to complete their education. If we replicate Panel B with experience defined as earnings year minus graduation year, the gap between the high reputation and low reputation profiles widens even more.

of the high reputation profile from the light grey dashed line that is parallel to the low reputation profile. The widening reputation gap is starkly different from the constant college wage premium observed in Panel A, and although the sample differs, it mirrors the results from column (A) of Table 1.

Our 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. This is consistent with the multidimensional nature of college quality, and with the fact that employers likely observe college identity, not 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.

The possibility that employers do not perfectly observe reputation makes the results in column (A) inconclusive as to the signaling role of reputation. At the end of this section, we consider a stronger signaling test that adds individual admission scores to the regression.

3.3.4. *Unconditional return to ability.* Column (B) of Table 1 estimates (6) including Icfes terms but not reputation terms, such that the coefficients represent the unconditional returns to ability, a^u (equation (3), 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 1 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 and Icfes increases as workers reveal their skill through their output. Column (B) 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.

This result 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 et al. (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.

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.

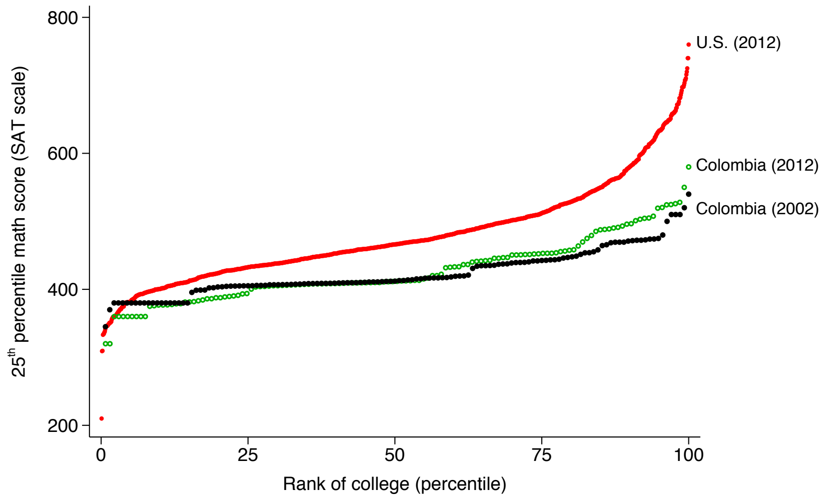


FIGURE 3. Ability sorting in Colombia and the U.S.

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.

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 3 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 Colombian college, with schools ranked on the x-axis according to this 25^{th} percentile.[23] 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 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 decline of a similar magnitude.

Despite these dynamics, by this measure Colombia's college market features substantially less sorting than that in the U.S. Figure 3 also shows the 25^{th} percentile SAT math scores for

[23] We plot the 25^{th} math percentiles for comparability with U.S. data. Other subjects and percentiles produce similar results.

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.[24]

A plausible explanation for the positive coefficient on the interaction of Icfes and experience in Table 1 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.[25]

3.3.5. *Conditional returns to ability and reputation.* Column (C) of Table 1 estimates equation (6) as written. In this joint specification, the coefficients reflect the conditional returns to reputation and to ability from equation (4) in Section 2. As Proposition 1 predicts, the coefficients on reputation and Icfes returns are lower than their respective unconditional returns in columns (A) and (B), but each is highly significant. Consistent with employer learning about ability, column (C) also shows a positive and significant coefficient on the interaction of Icfes and experience.

The main coefficient of interest is on the interaction of reputation with experience. Proposition 1 states that the conditional return to reputation should fall over time as the weight in wage setting shifts to the unobservable Icfes scores. This is similar to the Altonji and Pierret (2001) prediction for observable traits like race or schooling, but a unique feature of our setup makes our regression an even stronger test of the signaling role of reputation.

Specifically, in our joint specification one key characteristic—reputation—is a group-level mean of the other—Icfes. This implies that Icfes performance is a sufficient statistic for ability, α_i, and that *mechanically* the conditional reputation coefficients do not reflect the transmission of information on ability. This means that the return to reputation should fall with experience even if employers do not perfectly observe our measure of reputation; any learning about reputation will be reflected in the Icfes coefficients. Furthermore, unlike Altonji and Pierret (2001), our model predicts a negative coefficient on Reputation$\times t$ even

[24] 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 3 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.

[25] If we estimate Table 1 with Icfes scores normalized to mean zero and standard deviation one—as Arcidiacono et al. (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.

if there are interactions between ability, α_i, and human capital growth, h_{it}. These effects would also be captured by the Icfes$\times t$ term.

In sum, if college reputation serves purely as a signal of ability, Proposition 1 predicts a negative coefficient on the interaction of reputation and experience as the weight in wage determination shifts from the noisier characteristic, reputation, to the more precise characteristic, Icfes. Column (C) clearly rejects this. The reputation-experience interaction, although smaller in magnitude than in column (A), is still positive and significant.

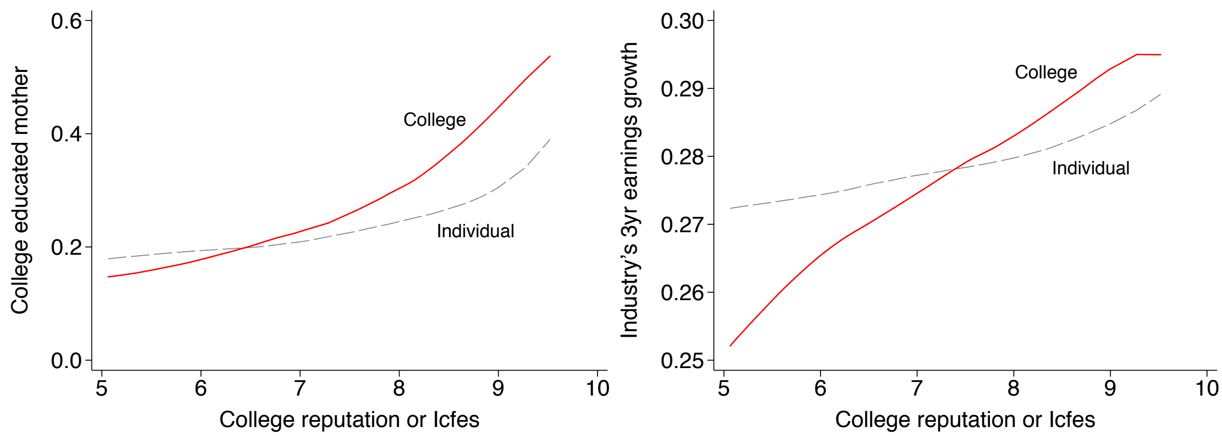
The increasing correlation of reputation and earnings in this joint regression is a descriptive result, but it is robust to a number of alternate specifications discussed in Appendix B.3. This result holds when we include controls for variation in earnings paths across gender, socioeconomic status, college programs, and regional markets. The returns to reputation increase with experience even when we allow earnings trajectories to vary with initial earnings, in the spirit of Farber and Gibbons (1996). Our finding is also unchanged when we use actual experience, defined by months of employment, rather than potential experience measured by graduation date, or when we restrict the sample to full-time employees with no work history prior to graduation.[26]

The rising return to reputation rejects a model in which reputation relates to wages only as a signal of ability. Instead, our finding suggests that attributes related to college membership other than ability influence earnings growth. These attributes could reflect sorting on other traits like socioeconomic status, or factors that contribute to skill acquisition while at school such as teaching or peer effects.

In our model, college membership attributes are denoted by v_{s_i}, and we suppose 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 4 provides suggestive evidence that both of these channels may be at work. First, Panel A considers socioeconomic status as measured by whether a student's mother has a college degree. The x-axis contains reputation when observations are colleges, and Icfes when observations are individuals (the possible values are the same). The solid line shows that as one moves from the college with the lowest reputation to that with the highest, the mean fraction of students with college-educated mothers increases from below 20 to above

[26] 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 et al. (2010).



Panel A. Socioeconomic status

Panel B. Industry's earnings growth

FIGURE 4. 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.

50 percent. The dashed line describes the individual-level relationship between students' Icfes scores and their mother's education, i.e., this is the the relationship that would exist if sorting into colleges were by Icfes only. Socioeconomic sorting is less pronounced in this hypothetical scenario than in the actual one. In other words, there is more sorting across colleges on mother's schooling than is predicted by Icfes scores alone.[27] This is consistent with a positive reputation premium ($v_1 > 0$), which could lead to a rising return to reputation if employers imperfectly observe both reputation and mother's education.[28]

Second, Panel B shows that the reputation premium, v_1, may be correlated with human capital growth. 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 from the Ministry of Social Protection, 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

[27] 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)—may account for some of the substantial role that socioeconomic status plays in college choice.

[28] Patterns similar to that in Panel A of Figure 4 emerge for traits related to family income, parents' occupation, and geography.

population-level relationship between industry earnings growth and Icfes scores. Graduates with 50^{th} percentile Icfes scores have 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. Panel B thus shows that graduates from higher-ranked colleges obtain jobs in industries with greater earnings growth, and this relationship holds even for students with similar ability. The increasing reputation coefficients in Table 1 may therefore reflect a career effect (Topel and Ward, 1992) in which better college reputation allows some individuals to be 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., Kaufman et al., 2013; Zimmerman, 2013a) that ultimately make individuals more productive.[29]

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. But we draw two conclusions from this descriptive result. First, the widening of earnings profiles across Colombian colleges is starkly different from the parallel nature of earnings profiles across schooling levels. Students may therefore suspect that their choice of college quality matters for their earnings trajectories in a way that their choice of educational quantity might not.

Second, the increasing correlation between reputation and earnings complicates an analysis of whether college reputation transmits information on ability. In considering whether years of schooling play such a role, Farber and Gibbons (1996) and Altonji and Pierret (2001) discuss how a correlation between years of schooling and subsequent human capital growth would similarly complicate matters. But they find no evidence that the returns to schooling increase with experience. This allows them to argue against the human capital hypothesis and in favor of an employer learning/statistical discrimination model in which schooling signals ability to the labor market. Our results so far reject a model in which reputation serves *solely* as a signal of ability but cannot say whether reputation serves *partially* as a signal of ability.

[29] 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).

To further explore the signaling role of reputation, we turn to an analysis of the introduction of a new signal of skill. The next section asks whether the creation of a field-specific college exit exam affected the conditional returns to college reputation and ability, and thus whether at least part of the return to reputation is informational.

4. The college exit exam

In this section we analyze the impact of the introduction of a college exit exam on the relative returns to college reputation and ability. We first describe the test, and then discuss our sample, empirical specifications, and results.

4.1. **The exit exam.** In 2004 the agency that administers the Icfes admission exam began another major initiative by introducing field-specific college *graduation* exams.[30] The exit exams are standardized and administered in every college that offers a related program. The fields range from relatively academic in orientation (e.g., economics and physics) to relatively professional (e.g., nursing and occupational therapy). The creation of these tests was a major undertaking, as it required coordination among departments in multiple colleges.

The stated intent of this effort was to introduce elements of accountability into the college system. There was a perception that some colleges, particularly recently created private institutions, delivered low quality instruction that information on skills-related outcomes could help to expose and correct. Consistent with this, school-level aggregate scores were made available and used by news outlets as part of college rankings.

Rather than focusing on its accountability dimension, we analyze the exit exam as potentially affecting students' capacity to signal their ability. This is consistent with anecdotal evidence that many students list exit exam scores on their CVs or on online profiles in job search websites. The exit exam may also have affected faculty recommendations or students' search behavior after learning their position in the national distribution of exam takers.

Each of these channels suggests that the exit exams increased the precision of employers' initial information about job applicants. Below we test how such an each increase in precision affects the relative returns to college reputation and ability.

4.2. **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. Columns (A) and (B) in Table 2 list the 55 field exams that were introduced between 2004 and 2007. Fields such as economics, engineering, and business were implemented first because in addition to large enrollments, they had long-standing traditions in Colombian higher education, and were thus offered by a high proportion of colleges. By 2007 all 55 fields were available, although none were required.

[30] These tests were initially labeled Ecaes, which stands for *Exámenes de Calidad de Educación Superior,* i.e., higher education quality exams. They are now called *Saber Pro*.

In 2009 the exams became mandatory, and a "generic competency" (*competencias genéricas*) exam was made available for programs without a corresponding field.

Although the exit exams were obviously field-specific, during the period we study there was no formal system assigning college majors to exam fields. This match is a necessary ingredient to determining which majors were treated. We therefore perform this assignment ourselves, using two approaches that produce very similar results. First, 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 2009.[31] Column (C) in Table 2 shows the resulting match of programs and exit exam fields.

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.[32] We use the name-matching procedure for our main results because students' exam choices are potentially endogenous, though our results are nearly identical under either method.[33]

[31] 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.

[32] 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*). 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*). Mining and metallurgical engineering is the only one of these four programs that appears in our final sample.

[33] Column (C) in Appendix Table B7 shows our main results when we define treatment by this exam-choice procedure. Column (D) shows our main results using a third procedure for matching programs to fields. In 2011, the Colombian Institute for Educational Evaluation began assigning programs to "reference groups" and requiring each group to take different exit exam components. We obtained these reference groups for the 2013 exam, but this test is significantly different from the 2004–2009 tests covered in Table 2—it contains

Table 2. 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	✓
	Zootecnia	Zootecnia	Agronomy	1,144	1	
	Ingeniería agronómica y agronomía	Agronomía	Agronomy	84		
	Administración	Administración	Business	28,406	46	✓
	Contaduría	Contaduría pública	Business	15,712	36	✓
	Economía	Economía	Business	8,646	21	✓
	Licenciatura exams (seven in total)	Educación	Education	16,910	21	✓
	Ingeniería industrial	Ingeniería industrial y afines	Engineering	12,331	25	✓
	Ingeniería de sistemas	Ingeniería de sistemas, telematica y afines	Engineering	11,312	25	✓
	Ingeniería civil	Ingeniería civil y afines	Engineering	7,347	19	✓
	Ingeniería electrónica	Ingeniería electrónica, telecomunicaciones y afines	Engineering	7,385	14	✓
	Arquitectura	Arquitectura y afines	Engineering	4,400	12	✓
	Ingeniería mecánica	Ingeniería mecánica y afines	Engineering	4,639	9	✓
	Ingeniería ambiental	Ingeniería ambiental, sanitaria y afines	Engineering	3,804	8	✓
	Ingeniería de alimentos	Ingeniería agroindustrial, alimentos y afines	Engineering	1,443	5	✓
	Ingeniería química	Ingeniería química y afines	Engineering	3,439	4	✓
	Ingeniería eléctrica	Ingeniería eléctrica y afines	Engineering	1,490	3	✓
	Ingeniería agronómica y agronomía	Ingeniería agronómica, pecuaria y afines	Engineering	1,474	3	✓
	Ingeniería agrícola	Ingeniería agrícola, forestal y afines	Engineering	903	1	
	Enfermería	Enfermería	Health	7,927	19	✓
	Medicina	Medicina	Health	7,767	8	✓
	Fisioterapia	Terapias	Health	5,126	8	✓
	Odontología	Odontología	Health	2,616	7	✓
	Bacteriología	Bacteriología	Health	2,211	6	✓
	Nutrición y dietética	Nutrición y dietética	Health	1,019	3	✓
	Optometría	Optometría, otros programas de ciencias de la salud	Health	629	3	✓
	Psicología	Psicología	Social sciences	11,726	24	✓
	Derecho	Derecho y afines	Social sciences	15,934	21	✓
	Comunicación e información	Comunicación social, periodismo y afines	Social sciences	6,441	16	✓
	Trabajo social	Sociología, trabajo social y afines	Social sciences	4,201	7	✓
2005	Biología	Biología, microbiología y afines	Natural sciences	3,257	5	✓
	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	✓
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	✓
		Ingeniería de minas, metalurgia y afines	Engineering	1,554	2	✓
		Otras ingenierías	Engineering	720	2	✓
		Ingeniería biomédica y afines	Engineering	358	1	
		Diseño	Fine arts	4,609	7	✓
		Publicidad y afines	Fine arts	1,320	5	✓
		Artes plásticas, visuales y afines	Fine arts	2234	4	✓
		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	✓
		Lenguas modernas, literatura, lingüística y afines	Social sciences	841	4	✓
		Antropología, artes liberales	Social sciences	668	3	✓
		Geografía, historia	Social sciences	647	2	✓
		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 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 31. 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. Checkmarks in column (G) indicate programs included in our final sample. See the text for details on trimming, balancing, and selecting programs.

We then define a binary treatment variable δ_{pc} equal to one for students in program p and graduation cohort c that had an available exit exam in their 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.[34] For example, $\delta_{pc} = 1$ for psychology students who graduated in 2005 or later because the psychology field 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. We will often refer to "program groups" defined by the introduction year of their assigned exam field. For example, "2004 programs" are those with an exam field that appeared in 2004, while "2009 programs" had no field until the introduction of the generic exam in 2009.

Figure 5 describes our "first stage," showing that the introduction of exit exam fields led to sharp increases in the fraction of students taking the test. All program groups exhibit large increases in test taking in the cohorts one year after exam introduction. 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.[35] Thus, the introduction of fields led to substantial increases in test taking rates, although these are not equal to 100 percent. This reflects that until 2009 students were not required to take any exam and in fact could take any test they wished, although in practice 94 percent of all test takers in our sample took the field we assign to their program.[36]

To summarize, we define a treatment indicator, δ_{pc}, at the program-cohort rather than at the individual level, i.e., we define students as treated if they were near graduation when a field exam appeared in a subject related to their major. 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 are under 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 available in that individual's labor market, we do not know precisely how it affected what firms observed about her.

numerous subject-specific modules and several common components. Our results are qualitatively similar when we use the 17 large college-level reference groups to define programs, but we prefer the Ministry of Education's programs because they align better with the granularity of the 2004–2009 exam fields.

[34] 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.

[35] 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.

[36] In addition, students could repeat exams; less than three percent of all test takers did so.

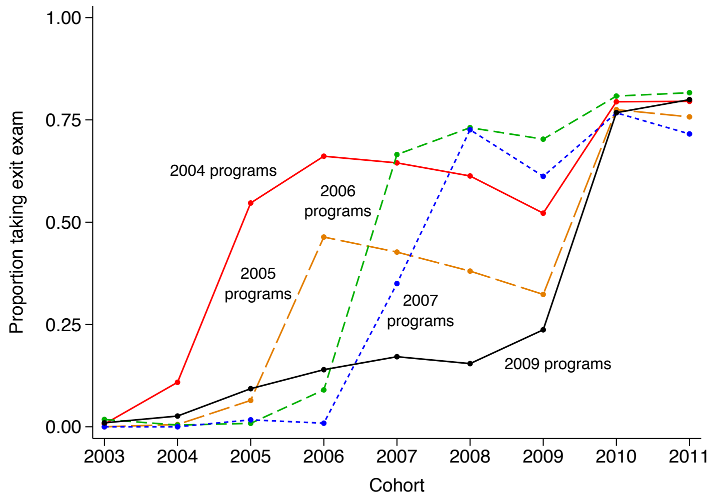


FIGURE 5. Proportion of students taking exit exam by program group

Notes: Lines represent program groups defined by the year in which their assigned exit exam field was introduced (see Table 2). The figure includes 2003–2011 graduates from all programs in our data, even those excluded from our main analysis sample for reasons described below.

4.3. **Sample.** In this section we describe how we select the cohorts, programs, and colleges we include in our empirical analysis.[37]

4.3.1. *Cohorts.* 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.[38]

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. 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 empirical section describes how we address this data constraint.

[37] We emphasize that the sample we use for the exit exam analysis differs from that in Section 3. In Section 3 we study earnings growth, which requires a sample of recent graduates (for whom we observe initial earnings) that remain in the labor market for four years. In this section we explore how the exit exam affects wage determination at any level of experience. This sample is larger because it includes older, pre-exam cohorts and does not condition on initial labor market attachment. However, as we describe below, the sample includes a smaller number of programs and colleges in which we can cleanly identify effects of the exit exam.

[38] 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.

4.3.2. *Programs and colleges.* 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 2 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 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.[39] 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.[40]

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.[41] 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 2 shows the number of colleges that offer each program after trimming and balancing the sample. We exclude any program offered at a single school.[42] The final sample includes the 39 programs with checkmarks in column (G) and any colleges that offer those programs after trimming and balancing. This covers

[39] As stated above 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.

[40] Students do not appear in our earnings records if they are not formally employed. This could be because they are unemployed, in graduate school, out of the labor force, or working in the informal sector. The fact that we observe earnings conditional on formal employment raises a sample selection issue. Below we present evidence that the exit exam had little impact on the likelihood of formal employment.

[41] Columns (B)-(D) of Appendix Table B6 show how our main results vary by the percentile below which we drop small school-program-cohort cells. 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.

[42] Columns (E) and (F) in Appendix Table B6 show how our main results vary when we include only programs offered at more than two colleges. This gives a more precise estimate of the return to reputation within each program, but it reduces the variation in our treatment variable δ_{pc}. In general, our results are not sensitive to whether we require programs to be offered by two, three, or four or more colleges.

Table 3. Summary statistics for exit exam sample

	Year program received exit exam (program group)				
Variable	2004	2005	2006	2009	All
# graduates in 2003–2009	131,962	2,014	1,043	11,033	146,052
# earnings observations	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 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.

146,052 graduates from 94 colleges. We observe four years of earnings per student on average, resulting in 581,802 total observations.

4.3.3. *Descriptive statistics.* Table 3 displays summary statistics for the final sample.[43] We present statistics separately for program groups defined by the year each program received its assigned exit exam field. All 94 colleges in the sample offer at least one program with a 2004 exam field. Less than ten percent of students graduate from one of the twelve post-2004 programs, and only 25 schools offer one or more of these programs. In particular, we assign only one program to each of the 2005 and 2006 exit exam years, so our identification mainly comes from the comparison of 2004 and 2009 programs.

Table 3 also reports students' mean Icfes scores and the average reputation of the college from which they graduate. As in Section 3 we define τ_i as student i's Icfes percentile divided

[43] Appendix Table B4 presents analogous statistics for students from excluded colleges/programs and those we dropped due to missing values. On average these excluded students have only slightly lower Icfes scores but attend colleges with reputations that are four percentile points lower. Their average return to reputation is about six percentage points lower, but they have a similar average return to Icfes.

by ten, so one unit represents ten percentage points in the full population of Icfes exam takers.[44] Since less than half of all high school graduates eventually enroll in college and, of those, about 50 percent graduate, the distribution of Icfes scores in our sample is right-skewed with mean around the 77^{th} percentile—or 7.7 points.

We use the same measure of college reputation as in Section 3; R_s is the average Icfes percentile for 2000–2003 exam takers who graduated from school s. We use these exam cohorts to avoid capturing any enrollment effects from the exit exam rollout.[45]

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. These metrics suggest that identification primarily arises from higher reputation schools, which tend to offer the less-common programs with delayed exit exam fields.

The last two rows in Table 3 report the conditional returns to reputation and ability (Icfes) within each program group. These are analogous to the r and a coefficients from equation (4) in Section 2, except that these are averages across the multiple years of experience we observe using 2008–2012 earnings.[46] In Table 3 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

[44] We note that the Icfes percentiles we use in this section are different from those in Section 3. In Section 3, 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 this section, 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 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.

[45] This means that our reputation measure is calculated using Icfes percentiles from the Colombian Institute for Educational Evaluation records, while the individual Icfes percentiles we use in the regressions below are based on data from the Ministry of Education (see footnote 44). Column (E) in Appendix Table B7 shows our main results are similar when we define reputation using Icfes percentiles from the Ministry of Education's records and only students in the exit exam sample. Columns (F) and (G) in the same table show that our results are also unchanged when we convert reputation and Icfes measures to $N(0,1)$ variables (consistent with the theory in Section 2), or when we define reputation at the school-program level rather than the school level.

[46] To be consistent with our empirical specifications below, the regressions that estimate the returns to reputation and ability in Table 3 also include program-cohort dummies and a quadratic in experience interacted with program dummies.

programs, but only a three percent increase for 2009 programs.[47] Conversely, 2009 programs have returns to Icfes that are almost twice as large as those in 2004 programs.

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 questions in our empirical analysis below.

4.4. **Empirical specifications and results.** In this section we describe and estimate a benchmark specification that tests the effects of the exit exam introduction on the returns to reputation and ability. We complement these results with three types of robustness checks. First, we add controls for potential experience and graduation cohort to address issues with our data structure and 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 use balance and placebo regressions to test for differential sorting or concurrent macroeconomic trends. We conclude with suggestive evidence on complementary effects of the exam introduction, including responses in student effort and enrollment decisions.

4.4.1. *Benchmark differences in differences specification.* To study the introduction of the exit exams, 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 if the impacts of college reputation and Icfes on earnings changed with the introduction of the exit exams. Our benchmark specification therefore relates changes in the returns to reputation and ability to the staggered rollout of the exit exam fields.

Consider the regression:

$$w_{ipct} = d_{pc} + f_p(t) + r_{pc}R_{s_i} + a_{pc}\tau_i + e_{ipct}, \tag{7}$$

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. d_{pc} are dummies for program-cohort cells and $f_p(t)$ is a quadratic in experience interacted with program dummies. The key feature of this "first-step" specification is that it estimates conditional returns to college reputation, r_{pc}, and to ability, a_{pc}, separately for each program-cohort cell.

[47] The 13.3 percent average conditional return to reputation for all programs is higher than the 7.9 percent period-zero return in Column (C) of Table 1. This is due to the fact that we observe the 2003–2004 cohorts at higher experience levels, and, as indicated in Table 1, the return to reputation rises with experience. 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 technicians). For related issues see Hastings et al. (2013) and Urzua et al. (2015).

A second-step regression then 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:

$$(8) \qquad \hat{r}_{pc} = \mu_p + \mu_c + \beta^r \delta_{pc} + \upsilon_{pc},$$

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 similar two-step procedure, where the second step is a regression weighted by the standard errors from the first step. We instead use a single-step specification to identify changes in the *relative* weights of college reputation and Icfes on earnings. Plugging (8) and a similar equation for $\hat{a}_{pc}$ into (7) yields our benchmark specification:

$$(9) \qquad w_{ipct} = d_{pc} + f_p(t) + (\mu_p + \mu_c + \beta^r \delta_{pc}) R_{s_i} + (\nu_p + \nu_c + \beta^a \delta_{pc}) \tau_i + e_{ipct}.$$

Equation (9) is analogous to regression (5) from Section 2, except we use differences in differences variation in treatment. It controls for program-specific experience effects as well as level differences in daily earnings across program-cohort cells. It also 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}.

Our main prediction from Proposition 2 is $\beta^r < 0$ 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 or an effect on recommendation letters. 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 in setting wages, the return to reputation falls ($\beta^r < 0$) and the return to ability rises ($\beta^a > 0$).[48]

[48] 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. This is important because 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. Furthermore, Appendix A.4 shows that the exit exams should have no effect on the *unconditional* return to reputation (i.e., the coefficient from a regression that includes only reputation) and a positive but smaller effect on the unconditional return to Icfes. Appendix Table B9 presents results consistent with this prediction; in programs with access to exit exams, the unconditional return to reputation declines and the unconditional return to ability increases, but both effects are smaller and statistically insignificant.

Column (A) of Table 4 estimates benchmark specification (9). Like all other columns in Table 4 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 a test, students exposed to the exit exams see their daily earnings become more correlated with incoming collegiate ability and less correlated with college reputation. We can compare these coefficients to the mean returns across all programs in Table 3. The reputation effect is a bit less than one third of the mean return to reputation, and the Icfes coefficient is slightly more than one half of the mean return to Icfes.[49]

Figure 6 illustrates the 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 (see Table 2) and the black lines contain programs that did not receive an exam field until the 2009 generic exam. 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. This is consistent with a decline in the return to reputation in programs with access to the exit exam, and no change in this return for non-exit exam programs.

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.

There are two sources of caution in interpreting the results from (9)—one related to data constraints and one related to identification. The first arises because our data cover only seven cohorts 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 3 and from the different pre-exit exam slopes in Figure 6. We now describe robustness checks that address these two issues.

4.4.2. *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 first-period

[49] Appendix Table B5 presents the underlying returns to reputation and Icfes for each program and cohort group from the first-step equation (7). Averaging these returns with the appropriate weights yields estimates similar to those in column (A) of Table 4.

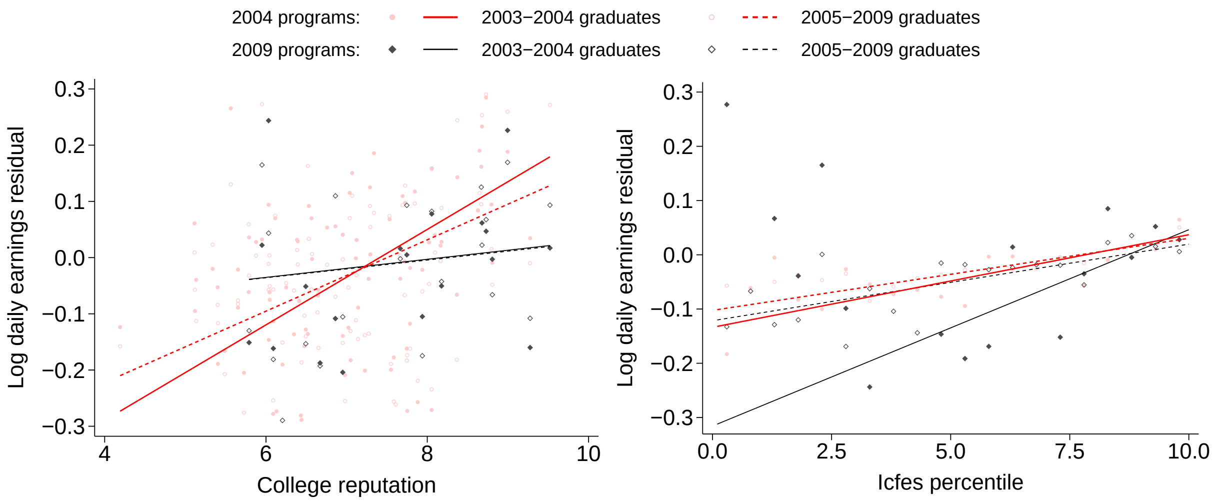
Table 4. Exit exam effects on returns to reputation and ability
Dependent variable: log average daily earnings

	(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 $\times$ δ_{pc}	−0.041**	−0.033**	−0.034	−0.046*	−0.026***	−0.051***
	(0.017)	(0.015)	(0.028)	(0.026)	(0.010)	(0.017)
Icfes $\times$ δ_{pc}	0.017***	0.018**	0.012	0.038***	0.016***	0.017***
	(0.006)	(0.007)	(0.009)	(0.010)	(0.005)	(0.005)
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.

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



Panel A. Return to reputation

Panel B. Return to ability

FIGURE 6. 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.

The sample for both panels omits 2005 and 2006 programs.

earnings for cohorts prior to the exam introduction (2003–2004). Note that our benchmark results are based on returns to reputation and ability in equation (9) that average across levels of potential experience, t.

Our data structure raises concerns if the returns to reputation and Icfes increase with experience, as suggested in Section 3. This could generate spurious results 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—for instance, top firms that hire from these programs may be more likely to supply early training activities. This could mechanically generate a $\beta^r < 0$ estimate since the post-exam cohorts (2005–2009) have lower potential experience than the cohorts that graduated before the exit exam appeared (2003–2004).

To address this issue we add further controls for experience to the benchmark specification. To illustrate, suppose we estimated equation (9) using only earnings at five years of potential experience, thus ensuring that we are comparing exposed and un-exposed 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 in our data is between four (using 2004–2008 graduates) and seven (using 2003–2005 graduates) years of experience.[50]

This procedure would yield four college reputation treatment effects (and four Icfes treatment effects), one for each year of potential experience. To combine these into a single estimate we modify equation (9) by removing the experience quadratics, restricting observations to those between four and seven years of experience, and fully interacting all fixed effects with experience dummies. This yields

$$w_{ipct} = d_{pct} + (\mu_{pt} + \mu_{ct} + \beta^r \delta_{pc}) R_{s_i} + (\nu_{pt} + \nu_{ct} + \beta^a \delta_{pc}) \tau_i + e_{ipct}, \tag{10}$$

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, respectively. If we were to allow the coefficient on $\delta_{pc} R_s$ to vary with experience, we would recover the four treatment effects from regressions that estimate equation (9) separately for each experience level.[51] The coefficients β^r and β^a are thus averages of the individual estimates, and they are identified only off variation within experience levels.

In short, if program variation in the interaction of reputation and experience mechanically biases our estimate of β^r downward, including these experience controls should move this coefficient toward zero. Column (B) in Table 4 shows the results from this specification. The inclusion of additional experience controls decreases the magnitude of the reputation effect only slightly. The lower magnitude suggests that returns to reputation do in fact rise more quickly in 2004 programs, although we cannot distinguish between inherent program differences and any effects of the exit exam on the interaction of reputation and experience. In either case, program differences in the returns to experience do not appear to be fully driving the reduction in the return to reputation. This is also true for the return to Icfes, as 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 Icfes ($\nu_p t \tau_i$ and $\nu_p c \tau_i$) to the benchmark specification (9).[52]

---

[50] In principle, we can identify treatment effects at three years of experience (using 2005–2009 graduates) and at two years of experience (using 2006–2009 graduates) because there are two programs in our sample that 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 leave little variation in our treatment variable δ_{pc} and produce noisy estimates.

[51] Appendix Table B8 contains the individual experience level estimates.

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

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

Including experience trends alone yields similar estimates to those from specification (10) since we limit the sample to earnings between four and seven years of experience. Adding 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.[53]

The results including linear trends appear in column (C) of Table 4. 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. Linear trends absorb much of the identifying variation because our sample includes only seven cohorts. This loss in precision suggests the effects of exit exam were not immediate but rather materialized over several years—an intuitive result if students gradually became more likely to report their scores.

4.4.3. *Restriction to similar programs.* The identifying assumption in our differences in differences approach is parallel trends in the returns to reputation and ability—i.e., that the correlations of reputation and Icfes with log earnings would have evolved similarly across programs in the absence of the exit exams. A potential violation of this assumption arises because 2004 programs display higher returns to college reputation than programs with delayed exam fields (see Table 3).

To address this we restrict identification to programs that are more similar. We define program groups G and supplement equation (9) with dummies for group-cohort cells interacted with reputation and Icfes (e.g., $\mu_{Gc}R_s$ and $\nu_{Gc}\tau_i$).[54] With these controls, the coefficients β^r and β^a are only identified by variation in exposure to the exit exam within groups of programs that have common characteristics.

Columns (D)-(F) in Table 4 show results using three types of program groups G. In column (D) we define program groups as the Ministry of Education's broader categorization of programs into eight areas (Table 2, columns (C) and (D)). Only two of these areas—social sciences and engineering—have multiple programs in different exam year groups; we therefore limit the sample to only these two.[55] The reputation effect in column (D) is similar

[53] 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).

[54] The full specification with program group controls is:

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

[55] The health program area also includes a single program with a delayed exit exam field (surgical instrumentation in 2006). Estimates analogous to column (D) that include the health program area yield coefficients

in magnitude to those in previous columns, while the Icfes effect is more than double that in prior specifications. Both estimates are statistically significant at the ten percent level despite the fact that the program restriction reduces the number of schools from which we identify a return to reputation.[56]

In column (E) we define program groups by pre-exit exam returns to college reputation. We first estimate a conditional return to reputation for each of the 39 programs in our sample using only the 2003–2004 graduation cohorts (i.e., $\hat{r}_{p,2003-2004}$).[57] We then define program groups G by quartiles of these returns, with 9–10 programs per group. This definition 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.[58] The reputation effect in column (E) is smaller than in earlier specifications, consistent with some inflation in our estimates due to differences in pre-treatment returns; 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 changes in the returns to reputation.

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 essentially unchanged from that in our benchmark regression.

4.4.4. *Checks for balance and placebo tests.* As a further robustness check, we run balance regressions to test if the exit exam rollout was correlated with changes in graduates' observable characteristics. For example, if high ability students switched into early exam programs to have access to the tests, or if low ability students switched out to avoid them, we would expect to see differential changes in average Icfes scores across programs. Similarly, any effects of the exit exam on school choice should appear as changes in average reputation across programs. Appendix Table B10 shows little evidence of these behavioral responses using standard differences in differences regressions; the changes in Icfes and reputation measures are less than one percentile larger in early exam programs, and are insignificant. These results likely reflect high costs to switching programs in Colombia and the fact that

with similar magnitudes, but they are not significant at the ten percent level because identification in the health program area comes from this single program.

[56] We note, however, that column (D) of Table 4 does not adjust standard errors to account for the reduced number of program clusters, which is well below the rule of thumb suggested by Angrist and Pischke (2009).

[57] We do this by estimating equation (7) using only the 2003–2004 cohorts and replacing the r_{pc} and a_{pc} coefficients with r_p and a_p.

[58] Appendix Table B5 presents these program-specific returns to reputation (and returns to ability). Many of the 2009 programs have low returns to reputation (e.g., anthropology) and are thus matched to low-return 2004 programs (e.g., nutrition and dietetics). There are also high-return 2004 and 2009 programs that are matched in this specification (e.g., civil engineering and administrative engineering).

our sample predominantly includes students who enrolled prior to the existence of any exit exams.[59] They also support our identifying assumption of parallel trends; one might expect to see behavioral responses to differential macroeconomic trends that are correlated with the introduction of the exams.

Appendix Table B10 also shows little evidence that the exit exam affected the probability of formal employment—a potential sample selection concern since we do not observe earnings for non-employed or informal workers. The differences in differences estimate suggests that formal employment increased 1.7 percentage points more in programs with exit exam fields, but this effect is not statistically significant and is small relative to the mean of 65 percent.

A separate placebo test in Appendix Section B.11 replicates our main results in Table 4 using college *drop-outs* rather than college graduates.[60] College drop-outs are a compelling placebo group because they enroll in the same colleges and programs as graduates but are significantly less likely to have ever taken an exit exam. Fewer than 20 percent of drop-outs in our sample ever took any exit exam, and there is almost no change in the proportion taking the exam across the 2003–2009 drop-out cohorts. Conversely, more than 50 percent of college graduates in our main sample took the exam, and there are sharp increases in test-taking with the exit exam rollout (see Figure 5).

Appendix Table B11 shows coefficients for drop-outs that are analogous to column (A) of Table 4. There is little evidence that changes in drop-outs' returns to reputation and ability are correlated with the staggered introduction of the exit exam fields. If anything, the return to reputation for drop-outs increases with the exam rollout, although the effect is insignificant. The point estimate on the Icfes effect is close to zero.[61]

4.4.5. *Complementary effects of the exit exam.* To conclude this section, we present suggestive evidence on other outcomes that are consistent with a causal effect of the exit exams on wage determination. Column (A) in Table 5 shows how the exit exams affected graduation timing. 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 the exam increased the duration of students' college careers; individuals in programs with exit exam fields took about one quarter of a year longer to graduate.[62] This result is consistent with increased student effort in response to the exit exam, or with colleges taking

[59] Colombian colleges do not make it easy for students to change majors. Anecdotally, switching may require applying *de novo* and essentially forfeiting all previous coursework.

[60] For this test we include 2003–2009 drop-outs from the colleges and programs in our main sample. We use the same definitions of Icfes percentiles and college reputation, but we redefine potential experience as years since dropping out.

[61] For both the reputation and Icfes effects, the difference between the graduate and drop-out coefficients is marginally insignificant at the ten percent level.

[62] The mean time to graduation in our sample is 5.6 years with a standard deviation of 1.2 years.

Table 5. 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 3. We report the coefficient on our treatment variable, δ_{pc}. The regression 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 3. 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 3 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 and cohort dummies.

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

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

steps to prepare students for the test. There is anecdotal evidence of colleges seeking to influence their students' and by extension their own exit exam performance. The activities reported range from offering free of charge "boot camp" preparation sessions for test takers, to more overt "gaming" strategies such as excluding certain students.[63]

[63] These results suggest that graduation cohort may be endogenous in the estimation of our main treatment effects in Table 4. We address this concern by estimating specification (9) with cohorts redefined by *predicted* rather than actual graduation date, where predicted graduation is based on the year of enrollment. Because selective graduation also affects labor market experience, we replace our measure of potential experience with years since predicted graduation. Column (G) of Appendix Table B6 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.

Column (B) presents a similar differences in differences specification with log average daily earnings as the dependent variable.[64] Across the experience levels we observe (0–9 years), earnings increased seven percent more in programs with early exam fields. An increase in average earnings could have occurred if the exit exam improved match quality and hence raised 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.

The final result in Table 5 asks whether the exit exams altered individuals' school or program choices. If the exit exam altered the relationship between earnings and college reputation, one might expect to see enrollment decisions respond to schools' exit exam performance. This is consistent with the stated intent of the exams and with the public release of aggregate test results.

Column (C) explores how the ability of incoming student bodies changed as the exit exam revealed new information about college and program quality. For this regression, we define two measures of reputation using a population of college graduates who took both the admission and exit exams. We define Icfes reputation as mean Icfes percentile at the *school-program* level, i.e., programs within the same college vary in reputation. Similarly, exit exam reputation is the school-program mean exit exam percentile. We calculate these measures using only students who took the exit exam in 2009–2011, when it was required of all graduates. We convert Icfes and exit exam scores to percentiles within this population so that both reputation measures are on the same scale.[65]

Icfes and exit exam reputations are highly correlated but not perfectly so; some school-programs score better on the exit exam than their graduates' admission scores would predict, and others underperform. We suppose that the exit exam reputation contains new information on school-program quality, and that this information gradually became available to students entering college in 2003–2009 during the exam rollout. Starting with the 2005 enrollment cohort, students entering programs with early exam fields had access to this new information through published rankings of school-program performance. Information on exit exam performance was unavailable for programs with delayed fields until after 2009.

Column (C) in Table 5 shows how this gradual release of information relates to the Icfes scores of incoming cohorts. This regression is analogous to our benchmark specification (9) 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

[64] The regression in column (B) of Table 5 also includes a quadratic in potential experience interacted with program dummies. Repeating this estimation using levels instead of logs yields the same conclusion, which suggests that this effect is not driven by an increase in the variance of earnings.

[65] As above, one reputation unit represents ten percentile points in this distribution of exam takers.

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.[66] 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.

The results show that in programs where exit exams were introduced, 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 scores saw increases in the ability of their incoming classes, while the average ability of entrants declined in school-programs that underperformed in the exit exam. This suggests students selected different programs and/or colleges as new information on their quality became available through the exit exam. More broadly it is consistent with the hypothesis that students care about the informational content inherent in college identity.

In sum, we find evidence that 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 effects do not appear to be driven by data constraints or the non-random timing of the exam rollout by field. There is also suggestive evidence that the exit exam affected labor market matching as indicated by an increase in average earnings, and that it generated behavioral responses in the form of delayed graduation and preference for colleges and programs with better exit exam performance. Taken together, these results provide evidence that college reputation transmits information on individual ability to the labor market.

5. Conclusion

Debates like those surrounding affirmative action suggest that college has 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; Zimmerman 2013b) ask if college has a causal return, while others (e.g., Goldin and Katz, 2008) consider the evolution and determinants of the college wage premium—the average differential in earnings between college and high school graduates. Still other work explores the channels that may account for these findings. In seminal work, Spence (1973) suggests that a college premium can exist as a result of signaling and even if college has no value added.

We have explored analogous issues when the question is *which* college students attend rather than *whether* they attend. Specifically, we ask if students' demand for more prestigious colleges reflects a desire to transmit their ability to the labor market. This raises mechanisms that differ from those in Spence (1973). In that framework the key aspect is that schooling

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

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

is costly, but less so for individuals of high ability; in equilibrium, therefore, only the high ability choose to go to college. In contrast everyone in our data is a college graduate. In Spence (1973) there is no rationing of education; implicitly a single school sets a difficulty level and accepts anyone who wishes to attend. In our setting there are multiple colleges and many are selective. The question is thus one of educational quality rather than quantity: does college reputation transmit information on students with a common level of schooling?

Our contribution is to use administrative data on individuals' admission exam performance to calculate a measure of college reputation. We do so for an entire national market and link this measure to graduates' labor market outcomes. We incorporate this measure into the employer learning model (Farber and Gibbons, 1996; Altonji and Pierret, 2001), where reputation is defined using admission scores to nest a standard signaling model within a more general model that allows for a reputation premium reflecting other college membership attributes. This produces two findings.

First, if reputation has only a signaling function, then its effect should be fully reflected in graduates' initial earnings; we find reputation to be robustly correlated with earnings growth. Although this is a descriptive finding, it raises three points. First, it is consistent with the demand for more reputable schools being driven by career concerns in addition to a desire for higher starting wages. Second, it stands in contrast with results on the impact of years of schooling (e.g., Lemieux, 2006). Third, it leaves open the possibility that colleges produce value added—broadly understood—for students. For instance, colleges may allow them to benefit from peer effects, or to gain alumni networks. They may also allow students to sort on attributes not perfectly correlated with admission scores.

While our first result rejects that reputation has only a signaling function, our second shows that signaling does account for part of the earnings return to reputation. For this we study how the staggered introduction of an additional signal of skill—a field-specific college exit exam—affected the returns to reputation and ability. Consistent with predictions from our signaling framework, it lowered the former and increased the latter. In addition, we find suggestive evidence that the exit exam had other impacts: it may have raised average earnings by improving employer/employee matches, and it may have gradually begun to change school reputation itself.

These findings suggest that the population return to college reputation partially reflects signaling, but that demand for selective colleges may also arise from students' concern for subsequent wage growth. The bottom line is that characteristics of college systems matter for the "big sort" that occurs as students transition to the labor market via college. Some countries' college systems might facilitate the transmission of information on ability while others may promote the development of career networks. These traits may respond to policy or change endogenously over time, with implications for the distribution of opportunity.

References

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

Angrist, J., E. Bettinger, and M. Kremer (2006). Long-term consequences of secondary school vouchers: Evidence from administrative records in colombia. *American Economic Review.*

Angrist, J. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Arcidiacono, P., P. Bayer, and A. Hizmo (2010). Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics 2*(4), 76–104.

Biglaiser, G. (1993). Middlemen as experts. *The RAND Journal of Economics 24*(2), 212–223.

Bishop, J. (2004). Drinking from the fountain of knowledge: Student incentives to study and learn–externalities, information problems, and peer pressure. Working paper, Center for Advanced Human Resource Studies.

Bound, J., B. Herschbein, and B. T. Long (2009). Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives 23*(4), 119–146.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In E. Christofides and R. Swidinsky (Eds.), *Aspects of labor market behaviour: Essays in Honour of John Vanderkamp.* Toronto, Ontario: University of Toronto Press.

Card, D. and A. 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.

Card, D. and T. Lemieux (2001). Can falling supply explain the rising return to college for younger men? a cohort-based analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

Coate, S. and G. Loury (1993). Will affirmative-action policies eliminate negative stereotypes? *American Economic Review 85*(5), 1220–1240.

Dale, S. B. and A. B. Krueger (2002, November). 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, S. B. and A. B. Krueger (2014, November). Estimating the effects of college characteristics over the career using administrative earnings data. *The Journal of Human Resources 49*(2), 323–358.

Dang, H.-A. and H. Rogers (2008). The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? *The World Bank Research Observer 23*(2), 161–200.

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.

Epple, D. and R. E. Romano (1998, March). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review 88*(1), 33–62.

Epple, D., R. E. Romano, and H. Sieg (2003). Peer effects, financial aid and selection of students into colleges and universities: An empirical analysis. *Journal of Applied Econometrics 18*, 501–525.

Farber, H. S. and R. Gibbons (1996, November). Learning and wage dynamics. *Quarterly Journal of Economics 111*(4), 1007–47.

Goldin, C. and L. Katz (2008). *The race between education and technology.* Cambridge, Massachusetts: Harvard University Press.

Hastings, J., C. Neilson, and S. Zimmerman (2013). Are some degrees worth more than others? evidence from college admission cutoffs in chile. Mimeo, National Bureau of Economic Research Working Paper No. 19241.

Heisz, A. and P. Oreopoulos (2002). The importance of signaling in job placement and promotion. Technical report, Mimeo, University of California at Berkeley.

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

Hoxby, C. (2009). The changing selectivity of american colleges. *Journal of Economic Perspectives 23*(4), 95–118.

Hoxby, C. and C. Avery (2012). The missing 'one-offs': The hidden supply of high-achieving, low income students. Working Paper 18586, National Bureau of Economic Research.

Hoxby, C. and S. Turner (2013). Expanding college opportunities for high-achieving, low income students. Working Paper 12-014.

Hoxby, C. M. (1997). How the changing market structure of u.s. higher education explains college tuition. Technical report, National Bureau of Economic Research Working Paper No. 6323.

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

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kaufmann, K. M., M. Messner, and A. Solis (2013). Returns to elite higher education in the marriage market: Evidence from chile. Mimeo, Bocconi University.

Lemieux, T. (2006). The mincer equation thirty years after schooling, experience, and earnings. In S. Grossbard (Ed.), *Jacob Mincer, A pioneer of modern labor economics*, pp. 127–145. Springer Verlag.

MacLeod, W. B. and M. Urquiola (2013). Anti-lemons: Reputation and educational quality. Mimeo, Columbia University.

Mincer, J. (1974). *Schooling, experience and earnings.* New York: Columbia University Press.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy 78*(2), 311–329.

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

Saavedra, J. (2009). The learning and early labor market effects of college quality: A regression discontinuity analysis. Mimeo, Harvard University.

Saavedra, J. (2012). Resource constraints and educational attainment in developing countries: Colombia 1945-2005. *Journal of Development Economics 99*(1), 80–91.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 3*, 355–374.

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.

Urzua, S., J. Rodriguez, and L. Reyes (2015). Heterogenous economic returns to postsecondary degrees: Evidence from chile. Mimeo, University of Maryland.

Zimmerman, S. (2013a). Making top managers: The role of elite universities and elite peers. Mimeo, Yale University.

Zimmerman, S. (2013b). The returns to college admission for academically marginal students. Mimeo, Yale University.

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 1 and 2, which are the basis for our empirical results 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 suppose that we can define two measures of α_i in our data. 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 college reputation in the following sense:

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

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:[67]

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

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

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

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:

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

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:

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

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:

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

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)

[67] 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 $E\{\alpha_i\} = 0$.

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 our measure of college reputation, R_{s_i}. We make this assumption to derive benchmark predictions consistent with the literature on observable characteristics like years of schooling. Below we discuss the empirical implications if employers do not perfectly observe R_{s_i}, a possibility given its finely-grained nature and the ambiguities in defining college quality.

Our measure of reputation, R_s, captures the pre-college ability of individuals at college s, i.e., the quality of its "student inputs." In setting wages employers are rather interested in 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 $\rho^{\mathscr{R}}$ denotes the precision of $\mathscr{R}_s$.[68]

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 $E\left\{v_s | R_s\right\} = v_0 + v_1 R_s$, 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:

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

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:

$$y_i = \alpha_i + v_0 + v_1 R_{s_i} + \epsilon_i, \tag{A7}$$

[68] 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, 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 Armed Forces Qualification Test scores are unobserved.

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

$$y_{it} = \alpha_i + v_0 + v_1 R_{s_i} + \epsilon_{it}, \tag{A8}$$

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 $\rho^{\bar{y}}$.[69]

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

A.3. **Predictions for earnings growth.** 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. In Section 3 we estimate three 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 the relationship between the signals of individual ability and wages net of human capital growth, which yields the regressions:

$$\hat{w}_{it} = r_t^u R_{s_i} + e_{it}^R \tag{A11}$$

$$\hat{w}_{it} = a_t^u \tau_i + e_{it}^\tau \tag{A12}$$

$$\hat{w}_{it} = r_t R_{s_i} + a_t \tau_i + e_{it}, \tag{A13}$$

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

[69] The assumption that the precision of $\bar{y}_{it}$ is time stationary also follows Farber and Gibbons (1996).

(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):

$$\begin{aligned}\hat{w}_{it} &= \pi_t^{\mathscr{R}}\left(\pi^{\alpha|R}R_{s_i}+v_0+v_1R_{s_i}\right)+\pi_t^{y}\left(\alpha_i+v_0+v_1R_{s_i}+\epsilon_i\right)\\ &\quad+\left(1-\pi_t^{\mathscr{R}}-\pi_t^{y}\right)\left(\alpha_i+v_0+v_1R_{s_i}+\bar{\epsilon}_{i,t-1}\right)\\ &= v_0+v_1R_{s_i}+\pi_t^{\mathscr{R}}\pi^{\alpha|R}R_{s_i}+\left(1-\pi_t^{\mathscr{R}}\right)\alpha_i+\epsilon_{it}^{w}, \qquad \text{(A14)}\end{aligned}$$

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

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:

$$\begin{aligned}E\left\{\hat{w}_{it}|R_{s_i}\right\} &= v_0+v_1R_{s_i}+\pi_t^{\mathscr{R}}\pi^{\alpha|R}R_{s_i}+\left(1-\pi_t^{\mathscr{R}}\right)\pi^{\alpha|R}R_{s_i}\\ &= v_0+\left(v_1+\pi^{\alpha|R}\right)R_{s_i}. \qquad \text{(A15)}\end{aligned}$$

Regression (A12) is given by:

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

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):

$$\begin{aligned}E\left\{\hat{w}_{it}|R_{s_i},\tau_i\right\} &= v_0+v_1R_{s_i}+\pi_t^{\mathscr{R}}\pi^{\alpha|R}R_{s_i}+\left(1-\pi_t^{\mathscr{R}}\right)\pi^{\alpha|\tau}\tau_i\\ &= v_0+\left(v_1+\pi_t^{\mathscr{R}}\pi^{\alpha|R}\right)R_{s_i}+\left(\pi^{\alpha|\tau}-\pi_t^{\mathscr{R}}\pi^{\alpha|\tau}\right)\tau_i. \qquad \text{(A17)}\end{aligned}$$

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 coefficient values imply the following proposition:

Proposition 1. *If wages are set equal to expected skill given the available information (equation* (1)*), 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.*

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 test the predictions from Proposition 1 in Section 3.

A.4. **Predictions for the introduction of a college exit exam.** In Section 4 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:

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

where:

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

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

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).[70] This in turn implies:

Proposition 2. *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 (*$\beta_t^a > 0$*).*

We test Proposition 2 in Section 4.

[70] Note from (A18) that the theory predicts no effect of the exit exam on the *unconditional* return to reputation. From (A19) we get $a_t^{u,exit} - a_t^u = (\pi_t^{\mathscr{R}} - \pi_t^{\mathscr{R},exit})(\pi^{\alpha|\tau} - \pi^{\alpha|R}\pi^{R|\tau}) > 0$, so the exit exam should have a positive effect on the unconditional return to ability, but this should be smaller than the effect on the conditional return to ability.

B. Empirical Appendix

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

B.1. **Section 3 sample.** In Section 3, 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.

The columns of Table B1 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.[71] 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.[72]

Our sample for Section 3 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.2. **Return to years of schooling in Colombia.** Our main result from Section 3 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 2.

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

[71] We consider workers as having formal earnings if they have at least one monthly earnings observation in a given year.
[72] F-tests for each characteristic strongly reject the hypothesis of joint equality across the four columns.

Table B1. Transition from college to the labor market
2008–2009 college graduates

Variable	(A) Went to graduate school	(B) # years formally employed in the four years after graduation: Zero	(C) 1 to 3	(D) Four
# 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.

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.[73]

Table B2 shows how the return to years of schooling in Colombia changes with experience. The use of cross-sectional data differentiates Table B2 from the panel data results in Farber and Gibbons (1996), Altonji and Pierret (2001), and Table 1 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.[74] The results suggest that an additional year

[73] 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).
[74] Regressions in Table B2 also include controls for experience and survey date.

Table B2. 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 $\times t$	-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.

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 B2 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 2.

Columns (C)-(D) of Table B2 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.3. **Robustness of increasing return to reputation.** Table B3 documents the robustness of our main result from Section 3: the return to reputation—even conditional on Icfes scores—increases with experience (see column (C) of Table 1). 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 B1. 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 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 2) 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 B3 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

Table B3. 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 B1 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 (C) in Table 1. 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.

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

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 B3 is identical to column (A), but we replace all experience terms with *actual* experience, defined as the number of months of employment since graduation.[75] This alternate measure of experience 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 B3 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.

B.4. **Summary statistics for Section 4 excluded students.** In selecting a sample for the exit exam analysis of Section 4, we exclude students with missing Icfes scores or formal sector earnings, as well as graduates from small colleges or small programs (see Section 4.3 for details). Table B4 displays summary statistics for this excluded population. These are analogous to those for included students in Table 3. The excluded population is about 50 percent larger in size than the sample for Section 4, 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 points lower. Their average

[75] 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 et al. (2010).

Table B4. Summary statistics for Section 4 excluded students

	Year program received exit exam (program group)					
Variable	2004	2005	2006	2007	2009	All
# graduates in 2003–2009	183,206	7,042	1,240	622	29,364	221,474
# earnings observations	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 4 (i.e., those not included in Table 3). All variables are defined identically as in Table 3. 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.

return to reputation is about six percentage points lower, but they have a similar average return to Icfes.[76]

B.5. **Returns to reputation and ability by program-cohort.** Our regression analysis in Section 4 is derived from a two-step estimation procedure. The first step equation (7) estimates conditional returns to reputation and ability separately for each program and cohort. The second step equation (8) relates these returns to the availability of the exit exam, captured in our treatment variable, δ_{pc}. Our benchmark specification (9) 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 (7). 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 2). 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

[76] In most cases, sample sizes are large enough that we can reject equality of mean characteristics between included (Table 3) and excluded (Table B4) students.

Table B5. Returns to reputation and ability by program and cohort

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)
Exam					Return to reputation			Return to ability		
Year	Program area	Program	N	Colleges	2003-04	2005-09	Diff.	2003-04	2005-09	Diff.
	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
2004	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
	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
2009	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	$\boxed{-0.041}$	-0.018	0.003	$\boxed{0.021}$

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 2 and the text for details.

Columns (F) and (G) report conditional returns to reputation for each program from specification (7) 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.

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 3, 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 program groups. The bottom rows of Table B5 show the average pre- and post-exam returns to reputation and ability for 2004 and 2009 programs.[77] The boxed numbers report the 2×2 difference in differences estimates. For example, the return to reputation declined from 13.8 percent 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 4. The 2×2 estimate for the return to ability is 2.1 percent, which is also close to our benchmark result.

Table B5 also helps to explain the estimates in columns (E) and (F) of Table 4. 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.[78] 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. **Sensitivity of exit exam effects to sample selection.** Table B6 tests the sensitivity of our exit exam results to the sample selection procedure described in Section 4.3. Column (A) of this table reprints our benchmark results from column (A) of Table 4.

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).

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

[77] Averages are weighted by each coefficient's inverse squared standard error from the first-step regression.

[78] 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 4 are from a specification including only 2003–2004 cohorts. This has little effect on the returns displayed in Table B5.

TABLE B6. 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 4. 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 the text.

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 the text.

Column (G) addresses the possible endogeneity of graduation cohort discussed in footnote 63. 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 (9) 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 $c = \tilde{c}+5$.

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

methods follow as in Section 4.3. 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 the text. Our results are not sensitive to this choice.

Table 5 in Section 4 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.7. **Sensitivity of exit exam effects to variable definitions.** Table B7 tests the sensitivity of our exit exam results to the definition of four key variables: log average daily earnings, w_{it}, treatment by the exit exam, δ_{pc}, college reputation, R_s, and Icfes scores, τ_i. In column (A), we replicate our benchmark results from Table 4. The bottom two rows of Table B7 display the mean returns to reputation and ability estimated from the 2003–2004 cohorts. These mean returns vary with the sample and with the variable definitions, and they provide a benchmark for the treatment effects in each column.

Column (B) uses a different definition of the dependent variable, log average daily earnings. Our benchmark specification calculates earnings using the income base for pension contributions. In Column (B) we instead use the income base for health contributions. The two earnings measure are very highly correlated, and the results with the health contributions measure are nearly identical.

Table B7. Sensitivity of exit exam effects to variable definitions
Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
		Definition of earnings	Definition of treatment (δ_{pc})		Definition of reputation and Icfes		
	Benchmark specification	Income from health payments	Most frequent 2009 field	Reference groups for 2013 exam	Reputation defined in sample	Converted to $N(0,1)$ variables	School-prog. level reputation
Reputation × δ_{pc}	−0.041**	−0.041**	−0.040**	−0.026	−0.032*	−0.093***	−0.038*
	(0.017)	(0.016)	(0.017)	(0.020)	(0.017)	(0.027)	(0.022)
Icfes × δ_{pc}	0.017***	0.017***	0.017**	0.014***	0.016**	0.040*	0.019**
	(0.006)	(0.006)	(0.006)	(0.004)	(0.006)	(0.023)	(0.007)
N	581,802	580,047	581,802	681,077	581,802	581,802	581,802
R^2	0.258	0.263	0.258	0.234	0.254	0.257	0.258
# programs	39	39	39	17	39	39	39
Mean return to reputation	0.133	0.134	0.133	0.123	0.119	0.214	0.132
Mean return to Icfes	0.029	0.029	0.029	0.030	0.032	0.098	0.027

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.

The mean returns to reputation and to Icfes in the bottom two rows are estimated from the same specification using only pre-exit exam cohorts (2003–2004), omitting the interaction terms, and including only a single reputation and a single Icfes term.

Column (A) is identical to column (A) in Table 4. All other columns estimate this same specification with different variable definitions.

Column (B) uses the income base for health contributions (rather than pension contributions) to calculate log average daily earnings.

Columns (C) and (D) use different treatment variables δ_{pc}. Column (C) defines treatment using the most common exam field taken by students from each program in 2009 (see footnote 32). Column (D) defines treatment using the Colombian Institute for Educational Evaluation's 2013 "reference groups" (see footnote 33). The agency assigns programs from each college to one of 17 reference groups, which take different exam modules. 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 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 select the sample and estimate column (D) as in the text with reference groups as our program variable.

Columns (E)-(G) use different definitions of reputation, R_s, and Icfes, τ_i (see footnotes 44 and 45). Column (E) defines reputation as school mean Icfes percentile using Icfes scores from the Ministry of Education and only graduates in the sample for this regression. Column (F) defines reputation as in our benchmark procedure, but instead of percentiles we use Icfes raw scores converted to mean zero and standard deviation one within the population of exam takers in the same year. In column (F) we also convert Icfes to a standard normal scale; we assign each Icfes integer percentile to the mean value of a truncated $N(0,1)$ with truncation points defined by these integer percentiles. Column (G) defines reputation as in our benchmark procedure, but at the school-program level rather than the school level.

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

Columns (C)-(D) use two different definitions of our treatment variable, δ_{pc}. In our benchmark specification, we define students as treated if they are one year from graduation when an exit exam is introduced with a name similar to their program; see Section 4.2 for details on this name-matching definition of treatment. In column (C), we define treatment based on the most common exam students in each program took in 2009, when all exit exam fields and the generic exam were available (see footnote 32 for details). 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 (C) are thus similar to our benchmark specification.

Column (D) uses a third procedure for matching programs to fields. 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. We obtained these reference groups for the 2013 exam, but this test is significantly different from the 2004–2009 tests covered in our analysis—it contains numerous subject-specific modules and several common components. The notes to Table B7 describe how we match the 2013 exit exam fields to the 2004–2009 fields. We then select the sample and estimate column (D) following all procedures in the text with reference groups as our program variable. Our results are qualitatively similar when we use the 17 reference groups to define programs, 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.

Columns (E)-(G) use different definitions of reputation and Icfes scores. Our main definition of reputation is based on Icfes scores from the agency that administers the test; we define reputation as school mean Icfes percentile using college graduates who took the exam in 2000–2003. Since many of the students in our exit exam sample took the Icfes before the period covered by our data from this testing agency, our main definition of Icfes scores uses integer percentiles from the Ministry of Education. Thus our reputation and Icfes measures are defined from different datasets, though the underlying Icfes scores we use are conceptually similar (see footnotes 44 and 45 for details).

In column (E), we define reputation as school mean Icfes using the Icfes percentiles from the Ministry of Education data. Furthermore, we use only students in the exit exam sample to calculate reputation. In this sample it is therefore precisely true, as in our theory, that reputation is just a noisy measure of Icfes. The results in column (E) are similar to our benchmark results in column (A).

In column (F), we convert both reputation and Icfes measures to normal variables with mean zero and standard deviation one. This is motivated by the normality assumption from our theoretical framework in Section 2. We calculate reputation using the same procedure as in our benchmark specification, but we normalize the underlying Icfes scores to mean zero and standard deviation one rather than to percentiles. To redefine– Icfes scores, we assign each Icfes integer percentile to the mean value of a truncated $N(0,1)$ with truncation points defined by these integer percentiles. The magnitudes of the estimates in column (F) are larger, but the effects are broadly similar in proportion to the mean returns to reputation and ability.

Column (G) is identical to column (A), but we calculate reputation at the school-program level rather than the school level. This is motivated by our empirical strategy, which relies on variation across programs in the returns to reputation. The point estimates are close to those in our benchmark specification, though the standard error on the reputation effect increases, likely because we use fewer students to calculate each reputation measure.

B.8. **Individual experience level exit exam effects.** As we discuss in Section 4.4.2, our benchmark specification (9) may lead to spurious estimates of the exit exam effects because we observe early and late cohorts at different levels of experience. To address this issue, we estimate specification (10), which adds experience controls to the benchmark regression. The resulting coefficients are weighted averages of the coefficients we would get from estimating (9) for each individual level of experience in 4–7—the years for which we can identify the exit exam effects.

Table B8 shows these individual experience level estimates. For reference, column (A) replicates the results from (10), which are identical to those in column (B) in Table 4. Columns (B)-(E) estimate equation (10) (or, equivalently, equation (9)) at each year of potential experience in 4–7. The estimates in column (A) are thus a weighted average of the estimates in columns (B)-(E). Our results are consistent across all experience levels, although in some cases we do not have enough power for statistical significance.

B.9. **Exit exam effects on unconditional returns.** Our benchmark specification estimates the effects of the exit exam on the *conditional* returns to reputation and ability. Our signaling model (Appendix A) also makes predictions for the *unconditional* returns, i.e., for the coefficients from regressions that include only reputation terms, or only Icfes terms.

The theory predicts no effect of the exit exam on the unconditional return to reputation. If employers perfectly observe our measure of reputation, new information from the exit exams merely confirms employers' expectations.[79] The theory predicts a positive effect on the unconditional return to ability because the exit exam reveals unobserved information that is also reflected in Icfes scores. This effect, however, should be smaller than the increase in the conditional return to Icfes.

[79] If employers do not perfectly observe our measure of reputation, then the prediction is similar to that for Icfes: the exit exam should have a small, positive effect on the unconditional return to reputation.

Table B8. Individual experience level exit exam effects
Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)
		Estimates by experience level, t			
	Within 4–7	$t = 4$	$t = 5$	$t = 6$	$t = 7$
Reputation × δ_{pc}	-0.033^{**}	-0.049^{**}	-0.020	-0.035^{**}	-0.032
	(0.015)	(0.020)	(0.014)	(0.016)	(0.035)
Icfes × δ_{pc}	0.018^{**}	0.021^{*}	0.012	0.014	0.027^{**}
	(0.007)	(0.010)	(0.007)	(0.012)	(0.012)
N	267,924	92,583	78,481	58,114	38,746
R^2	0.224	0.213	0.217	0.218	0.216
# programs	39	39	39	39	39
Cohorts	2003–2008	2004–2008	2003–2007	2003–2006	2003–2005

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.

Column (A) is identical to column (B) in Table 4. It estimates specification (10), which includes dummies for program-cohort-experience cells and interactions of both reputation and Icfes with program-experience and cohort-experience dummies. We restrict the exit exam sample to experience levels 4–7.

Columns (B)-(E) estimate specification (10) (or, equivalently, specification (9)), except we remove experience controls and run regressions separately for each experience level listed in the column header. Due to the timing of our earnings records, we can only include the cohorts listed in the bottom row.

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

Table B9 tests the effects of the exit exam on the unconditional returns. Column (A) reproduces our benchmark results from Table 4. Columns (B) and (C) present the unconditional effects. The coefficient in column (B) is from an estimation of our benchmark specification (9) excluding all Icfes terms, while column (C) presents the Icfes effect from the same regression excluding all reputation terms. The results are consistent with the theoretical predictions. Both unconditional effects have the same sign as the conditional effects, but they are smaller and statistically insignificant. This justifies our focus on the conditional returns to detect how the *relative* weight in wage setting shifts from college reputation to individual ability.

B.10. **Balance tests.** Section 4.4.4 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 B10 shows the results from these balance tests. These estimates are from simple differences in differences regressions that include program dummies, cohort dummies, and our indicator for exposure to the exit exams, δ_{pc}. The dependent variable for each regression is listed in the column header.

Table B9. Exit exam effects on unconditional returns
Dependent variable: log average daily earnings

	(A)	(B)	(C)
		Unconditional returns	
	Benchmark specification	Return to reputation	Return to ability
Reputation × δ_{pc}	-0.041^{**}	-0.029	
	(0.017)	(0.018)	
Icfes × δ_{pc}	0.017^{***}		0.005
	(0.006)		(0.007)
N	581,802	581,802	581,802
R^2	0.258	0.253	0.231
# programs	39	39	39

Notes: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. Column (A) of this table is identical to column (A) in Table 4. 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. Columns (B) and (C) use the same sample and specification, but column (B) excludes all Icfes terms from the regression, and column (C) excludes all reputation terms. Parentheses contain standard errors clustered at the program level.

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

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.

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.[80] The mean of this variable is 65 percent, and the remaining 35 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.

[80] 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.

Table B10. 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

Notes: All columns report coefficients on the treatment variable δ_{pc}. Parentheses contain standard errors clustered at the program level. We report the coefficient on our treatment variable, δ_{pc}.

The dependent variables in columns (A) and (B) are reputation and Icfes. The sample includes all students from Table 3. 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 3 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.

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

B.11. **Placebo test using college drop-outs.** Section 4.4.4 describes a placebo test that asks if the exit exam had a similar effect on college *drop-outs*. Table B11 depicts this test.

Columns (A) and (B) present "first-stage" regressions for the placebo test. The dependent variable is an indicator for taking the exit exam, and the specification is a simple differences in differences regression including program dummies and cohort dummies. We report only the coefficient on our treatment variable, δ_{pc}, which equals one if students had access to exit exam fields based on their program and cohort. The sample for column (A) includes the same 2003–2009 college graduates as in our main sample (see Table 3). Column (B) includes students from the same colleges and programs who dropped out in 2003–2009. In this regression, we use drop-out year to define cohorts and the treatment variable.

For college graduates, exposure to the exit exam is associated with a 50 percentage point increase in the likelihood of taking the exam; this mirrors the graphical evidence in Figure 5. For drop-outs, however, students' programs are not related to changes in exit exam taking. The proportion of drop-outs who took the exit exam increased by only 2.5 percentage points more in programs that received exam fields. These results suggest that college drop-outs are a compelling placebo group; they enroll in the same colleges and programs as graduates but exhibited little change in exam taking.

Table B11. 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

Notes: The sample for columns (A) and (C) includes college graduates and their earning observations (i.e., the same sample as in Table 3). 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 4. 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 date.

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

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

Columns (C) and (D) show the exit exam effects on the returns to reputation and ability. Column (C) replicates our benchmark results for graduates from Table 4. Column (D) estimates the same benchmark specification (9) using college drop-outs. There is little evidence that changes in drop-outs' returns to reputation and ability are correlated with the staggered introduction of the exit exam fields. The point estimate for drop-outs suggests a 1.1 percentage point *increase* in the return to reputation with the exam rollout, but this effect is insignificant. The point estimate on the Icfes effect is close to zero.[81]

To the extent that college drop-outs and graduates are subject to similar enrollment or macroeconomic trends, this placebo test supports the notion that our main results are attributable to the exit exams.

[81] For both the reputation and Icfes effects, the difference between the graduate and drop-out coefficients is marginally insignificant at the ten percent level.