Online Appendix for
The Big Sort: College Reputation and Labor Market Outcomes

W. Bentley MacLeod, Evan Riehl, Juan E. Saavedra, and Miguel Urquiola

A. Theoretical Appendix

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

A. Ability, admission scores, and college reputation

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

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

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

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

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

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

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

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

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

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:[44]

$$\text{(A3)} \qquad E\{\alpha_i|\tau_i\} = \frac{\rho^\tau}{\rho^\alpha + \rho^\tau}\tau_i = \pi^{\alpha|\tau}\tau_i$$

$$\text{(A4)} \qquad E\{\alpha_i|R_{s_i}\} = \frac{\rho^R}{\rho^\alpha + \rho^R}R_{s_i} = \pi^{\alpha|R}R_{s_i}$$

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

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 $\pi^{R|\tau} < 1$.

B. 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

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

the information, I_{it}, available regarding worker i in period t:

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

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

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

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

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

We suppose that employers' information set, I_{it}, includes college reputation, R_{s_i}.[45] While employers likely care about individuals' pre-college ability as captured by R_{s_i}, they also care about other attributes related to graduates' post-college skill. We therefore define a college's *labor market reputation* as the expected skill of its graduates: $\mathscr{R}_s = E\{\theta_i | i \in s\}$. It follows that $\theta_{i \in s} \sim N(\mathscr{R}_{s_i}, \frac{1}{\rho^{\mathscr{R}}})$, where $\rho^{\mathscr{R}}$ denotes the precision of $\mathscr{R}_s$.[46]

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:

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

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

[46] 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.

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

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

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

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

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

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

where ϵ_{it} includes human capital growth and other fluctuations in worker output. We suppose these are observed *after* setting wages in each period t, where $t = 0$ stands for the year of college graduation. We let $\bar{y}_{it} = \frac{1}{t+1}\sum_{k=0}^{t} y_{ik}$ denote mean worker output and suppose that the precision of y_{it} is time invariant and denoted by $\rho^{\bar{y}}$.[47]

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

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

where the weights on the signals are given by:

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

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

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

Equation (A9) describes employers' wage setting process given the information they observe, I_{it}. We do not observe I_{it}, and instead derive the implications of the wage equation for regressions on characteristics in our data. We use regressions that include controls for experience and graduation cohort to capture the time-varying effects (recall from above that $\hat{w}_{it} = w_{it} - h_{it}$). Here we focus upon the implications of the model for the relationship between the signals of individual ability and wages net of human capital growth. In particular, we consider three 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 (A12), a_t^u, is the unconditional *return to ability.* Specification (A13) estimates the conditional return to reputation, r_t, and the conditional return to ability, a_t.

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

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

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

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

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

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):

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

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

$$
\begin{aligned}
r_t^u &= v_1 + \pi^{\alpha|R} && \text{(A18)} \\
a_t^u &= v_1\pi^{R|\tau} + \pi^{\alpha|\tau} - \pi_t^{\mathscr{R}}\left(\pi^{\alpha|\tau} - \pi^{\alpha|R}\pi^{R|\tau}\right) && \text{(A19)} \\
r_t &= v_1 + \pi_t^{\mathscr{R}}\pi^{\alpha|R} && \text{(A20)} \\
a_t &= \pi^{\alpha|\tau} - \pi_t^{\mathscr{R}}\pi^{\alpha|\tau}. && \text{(A21)}
\end{aligned}
$$

These coefficients form the basis for Propositions 1 and 2.

D. Predictions for the introduction of a college exit exam

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

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

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

where:

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

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

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).[48] This in turn implies:

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

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

E. Predictions for earnings growth

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

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

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 $v_1, \pi^{\alpha|\tau}, \pi^{\alpha|R}, \pi^{R|\tau}, \pi_t^{\mathscr{R}} > 0$.

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

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

We take the predictions from Proposition 2 to the data in Section III and Appendix B.I.

F. Signaling vs. accountability effects of the exit exam

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

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

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

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

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

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

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

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

$$\begin{aligned} \beta_t^{u,r} &= r_t^{u,exit} - r_t^u \\ &= v_1^{exit} - v_1, \end{aligned} \tag{A27}$$

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

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

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

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

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

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

B. Empirical Appendix

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

A. Matching college programs to exit exam fields

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

Although the exit exams were field-specific, during the period we study there was no formal system assigning college majors to exam fields. This match is necessary to determine which majors were treated. We therefore perform this assignment ourselves, using three different approaches. In our benchmark approach, we consider all college majors belonging to the Ministry of Education's 54 core knowledge groups. These groups—which we label programs—aggregate approximately 2,000 college major names that vary across and within schools. For instance, the Ministry might combine a major named Business Administration at one college with one labeled Business Management at another if it considers that these have similar content. We assign each of the 54 programs to one of the 55 exam fields if one of the key words in the program name appears in the name of the field exam. We assign programs without any matching key words to the generic competency exam introduced in 2009.[49] Column (C) in Table B1 shows the resulting match of programs and exit exam fields. This is a more detailed version of the match displayed in Table 1 of Section II.

A second approach is to match programs to fields based on the most common exam students in each program took in 2009, when all fields and the generic exam were available. In this alternate procedure, we compute the percentage of 2009 test takers in each program that took a field exam introduced in 2004, 2005, 2006, or 2007, and the percentage that took the generic exam. We assign each program to an exit exam year using the maximum of these five percentages. This procedures differs from the name-matching method in only four programs: mathematics (*matemáticas, estadística y afines*), chemistry (*química y afines*), agricultural and forest engineering (*ingeniería agrícola, forestal y afines*), and mining and metallurgical engineering (*ingeniería de minas, metalurgia y afines*).[50]

[49] 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.

[50] This procedure matches mathematics and chemistry to the generic exam rather than the math-

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

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

Note: Columns (A) and (B) list exit exam fields and their year of introduction. *Licenciatura* includes seven exams covering pedagogical training intended for school teachers of preschool education, natural sciences, social sciences, humanities, math, French, and English. Column (C) shows the Ministry of Education's 54 core knowledge groups that we call programs. We match exam fields to programs using the method described in footnote 49. Thirteen fields did not match any program: 2004) Fonoaudiología, medicina veterinaria y zootecnia, terapia ocupacional; 2005) Ingeniería agroindustrial, ingeniería forestal, ingeniería de petróleos, técnico en electrónica y afines, técnico en sistemas y afines, tecnológico en electrónica y afines, tecnológico en sistemas y afines; 2006) Normalistas superiores, técnico profesional en administración y afines, tecnología en administración y afines. Column (D) shows eight program "areas" the Ministry of Education uses to categorize these 54 programs. Column (E) lists the number of 2003–2009 graduates with non-missing Icfes scores that appear in the earnings records. Column (F) reports the number of colleges offering each program after trimming and balancing the sample. A "Y" in column (G) indicates programs included in our final sample. See Appendix B.C for details on trimming, balancing, and selecting programs.

Mining and metallurgical engineering is the only one of these four programs that appears in our final sample.

In 2011, the agency that administers the exit exam began assigning programs from each college to one of 17 "reference groups," and they required each group to take different components. For a third procedure for matching programs to fields, we obtained these reference groups for the 2013 exam, but this test is significantly different from the 2004–2009 tests covered in Table B1; it contains numerous subject-specific modules and several common components. We assume reference groups that took the generic exam module in 2013 had no exit exam field for the 2003–2009 cohorts. We assume all other reference groups received an exit exam field starting with the 2005 cohort except for the natural sciences group, which received an exit exam field starting with the 2006 cohort. We then select the sample following all procedures described in Appendix B.C with reference groups as our program variable.

B. Sensitivity of exit exam effects to field-program matching

Table B2 tests the sensitivity of our exit exam results to the three field-program matching procedures described in Appendix B.A. In column (A), we replicate our benchmark results from Table 3, which matches exit exam fields to college programs based on their names.

In column (B), we match fields to programs based on the most common exam students in each program took in 2009. In our final sample, the assignment of programs to exit exam fields under this procedure differs from that in the name-matching method for only one program. The estimated effects in column (B) are thus similar to our benchmark specification. We use the name-matching procedure for our main results, however, because students' exam choices are potentially endogenous.

Column (C) uses the exam agency's 17 "reference groups" as our definition of programs. Our results are qualitatively similar under this procedure, though the reputation effect is smaller in magnitude with this coarser definition of treatment. We prefer using the Ministry of Education's programs to define treatment because they align better with the granularity of the 2004–2009 exam fields.

C. Section II sample

In this section we describe how we select the cohorts, programs, and colleges we include in our empirical analysis in Section II.

Our sample includes the 2003–2009 graduation cohorts. While our dataset covers students who enrolled in 1998–2012, there are few graduates before 2003

ematics and chemistry fields because the exit exam fields were less widely adopted in these programs. Agricultural and forest engineering is assigned to the 2005 exam group rather than the agricultural engineering field because 2009 test takers most commonly took the forest engineering field exam. Lastly, mining and metallurgical engineering is assigned to the 2005 exam group rather than the generic exam because students most commonly took the petroleum engineering field (*ingeniería de petróleos*).

Table B2—Sensitivity of exit exam effects to field-program matching

Dependent variable: log average daily earnings

	(A)	(B)	(C)
	Benchmark specification	Most frequent 2009 field	Reference groups for 2013 exam
Reputation × δ_{pc}	−0.041	−0.040	−0.026
	(0.017)	(0.017)	(0.020)
Icfes × δ_{pc}	0.017	0.017	0.014
	(0.006)	(0.006)	(0.004)
N	581,802	581,802	681,077
R^2	0.258	0.258	0.234
# programs	39	39	17

Note: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. All regressions include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. The sample for each regression includes experience 0–9. Parentheses contain standard errors clustered at the program level.
Column (A) is identical to column (A) in Table 3. All other columns estimate this same specification with different definitions of the treatment variable δ_{pc}. Column (B) defines treatment using the most common exam field taken by students from each program in 2009. Column (C) defines treatment using the Colombian Institute for Educational Evaluation's 2013 "reference groups." See Appendix B.A for details.

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.[51]

We define potential labor market experience, t, as calendar year minus graduation cohort, and drop any earnings observations prior to graduation. Our final sample therefore includes 2008–2012 earnings for 2003–2008 graduates and 2009–2012 earnings for 2009 graduates.

Two factors motivate how we select programs and colleges for our sample. First, our empirical specification will estimate the return to reputation for students in the same program and cohort. This return is imprecisely estimated when there are few students in the same school, program, and cohort, or when few colleges offer a given program. Second, our identification comes from the staggered introduction of the exit exam fields. Columns (D) and (E) in Table B1 show the Ministry of Education's categorization of programs into eight "program areas," and the number of 2003–2009 graduates in each program. Exam fields for most large programs in business, health, and engineering were introduced immediately in 2004. Field exams were delayed or never created for mostly smaller programs

[51] 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.

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.[52] Roughly 25 percent of these students never appear in our earnings records, and about 20 percent are missing Icfes scores or program variables. Excluding these leaves 225,856 graduates.

We then calculate the number of earnings observations across all experience levels for each school-program-cohort and drop cells below the 10^{th} percentile number of observations. After trimming, we drop school-programs with missing cohorts to balance the composition across all seven cohorts. Trimming eliminates ten percent of the sample with non-missing data and balancing the sample eliminates about 25 percent more. After this step, there are 147,788 graduates from 94 colleges.

Finally, in order to identify a return to college reputation, each program must be offered by at least two colleges. Column (F) in Table B1 shows the number of colleges that offer each program after trimming and balancing the sample. We exclude any program offered at a single school.

The final sample includes the 39 programs with a "Y" in column (G) and any colleges that offer those programs after trimming and balancing. This covers 146,052 graduates from 94 colleges. We observe four years of earnings per student on average, resulting in 581,802 total observations.

Table 2 in Section II displays summary statistics for the final sample. Table B3 here displays analogous summary statistics for students excluded in the process described above. The excluded population is about 50 percent larger in size than the sample for Section II, 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 return to reputation is about six percentage points lower, but they have a similar average return to Icfes.[53]

D. Sensitivity of exit exam effects to sample selection

Table B4 tests the sensitivity of our exit exam results to the sample selection procedure described in Appendix B.C. Column (A) of this table reprints our benchmark results from column (A) of Table 3.

In our benchmark sample, we calculate the number of observations in each school-program-cohort cell and exclude cells below the 10^{th} percentile. We exclude small school-program-cohorts because our empirical specification requires that we

[52] As stated we consider only graduating students who obtained 4–5 year degrees, the equivalent of bachelor degrees in the U.S. The sample for Section II begins with 136 colleges, but three of these only have 2010–2011 graduates in our records.

[53] In most cases, sample sizes are large enough that we can reject equality of mean characteristics between included (Table 2) and excluded (Table B3) students.

Table B3—Summary statistics for Section II excluded students

Variable	Year program received exit exam 2004	2005	2006	2007	2009	All
# graduates in 2003–2009	183,206	7,042	1,240	622	29,364	221,474
# earnings obs. in 2008–2012	440,635	18,648	2,747	1,808	74,090	537,928
# programs	30	5	1	1	18	55
# colleges	133	29	10	6	86	133
Reputation	6.97	8.25	6.33	6.59	7.63	7.09
	(1.21)	(1.08)	(0.87)	(0.66)	(1.11)	(1.23)
Icfes	7.52	9.03	6.18	6.20	7.80	7.61
	(2.39)	(1.32)	(2.45)	(2.34)	(2.19)	(2.35)
Log average daily earnings	10.83	10.96	10.62	10.33	10.76	10.82
	(0.67)	(0.72)	(0.57)	(0.45)	(0.71)	(0.68)
Return to reputation	0.080	0.040	0.060	1.393	0.041	0.075
	(0.021)	(0.055)	(0.033)	(0.121)	(0.032)	(0.017)
Return to ability	0.020	0.022	-0.020	-0.013	0.065	0.028
	(0.005)	(0.029)	(0.012)	(0.027)	(0.015)	(0.005)

Note: This table presents summary statistics for 2003–2009 graduates in our records that are excluded from the main analysis sample in Section II (i.e., those not included in Table 2). All variables are defined identically as in Table 2. Note that one reason we excluded these students is due to missing values on certain variables, so the statistics in this table are averages for only students who have values of each variable.

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 methods follow as in Appendix B.C. The signs are consistent across all trimming thresholds, though the reputation coefficient loses significance when we trim at the 25^{th} percentile, and the Icfes coefficient loses significance when we trim at the 5^{th} percentile or do not trim. The variation in statistical significance across trimming thresholds reflects the data demands of our empirical strategy, though the consistency of the signs is reassuring.

Columns (E) and (F) use different minimums for the number of schools that we require to offer each program. Our main specification in column (A) requires the bare minimum necessary to identify a return to reputation within each program: each program must be offered by two or more colleges. Columns (E) and (F) require that each program must be offered by three or more, and four or more, colleges. All other sample selection methods follow as in Appendix B.C. Our results are not sensitive to this choice.

Table 6 in Section II shows that the exit exam may have increased time to

Table B4—Sensitivity of exit exam effects to sample selection

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
		School-program-cohort trimming			# colleges in each program		
	Benchmark specification	No trimming	5^{th} percentile	25^{th} percentile	3 or more colleges	4 or more colleges	Predicted cohorts
Reputation × δ_{pc}	−0.041	−0.035	−0.040	−0.038	−0.042	−0.036	−0.044
	(0.017)	(0.015)	(0.016)	(0.031)	(0.018)	(0.017)	(0.018)
Icfes × δ_{pc}	0.017	0.006	0.012	0.020	0.016	0.015	0.017
	(0.006)	(0.007)	(0.008)	(0.009)	(0.006)	(0.006)	(0.010)
N	581,802	671,840	618,489	452,080	575,321	563,752	650,015
R^2	0.258	0.256	0.260	0.254	0.248	0.247	0.241
# programs	39	48	41	31	35	30	39
Trim percentile	10^{th}	0^{th}	5^{th}	25^{th}	10^{th}	10^{th}	10^{th}
Colleges/program	2+	2+	2+	2+	3+	4+	2+
Grad. cohorts	Actual	Actual	Actual	Actual	Actual	Actual	Predicted

Note: All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. All regressions include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. The sample for each regression includes experience 0–9. Parentheses contain standard errors clustered at the program level.
Column (A) is identical to column (A) in Table 3. All other columns estimate this same specification using different samples.
Columns (B)-(D) use different percentiles for the number of observations below which we drop small school-program-cohort cells. Our main specification in column (A) trims school-program-cohort cells below the 10^{th} percentile in terms of number of observations. Columns (B), (C), and (D) use no trimming, the 5^{th} percentile, and the 25^{th} percentile. In all cases we balance the sample after trimming so that each remaining school-program appears in all seven cohorts in our sample. All other sample selection methods follow as described in Appendix B.C.
Columns (E) and (F) use different minimums for the number of schools that we require to offer each program. Our main specification in column (A) requires the bare minimum necessary to identify a return to reputation within each program: each program must be offered by two or more colleges. Columns (E) and (F) require that each program must be offered by three or more, and four or more, colleges. All other sample selection methods follow as described in Appendix B.C.
Column (G) addresses the possible endogeneity of graduation cohort discussed in footnote 33. We create a new sample based on the year students entered college, $\tilde{c}$, rather than the year they graduated, c. Most university programs in Colombia have an official duration of ten semesters, so we define predicted graduation date as $\tilde{c}+5$. We include only students whom we predict to graduate in 2003–2009. In other words, this sample covers graduates who enrolled in 1998–2004, regardless of when they graduated. Because selective graduation also affects labor market experience, we replace our measure of potential experience with years since expected graduation, $\tilde{t} = y - (\tilde{c} + 5)$, where y is calendar year. We modify our benchmark specification (8) by replacing graduation cohort, c, with enrollment cohort, $\tilde{c}$, and potential experience, t, with predicted potential experience, $\tilde{t}$. We define the treatment variable $\delta_{p,\tilde{c}+5}$ as before with expected rather than actual graduation year—i.e., $\delta_{p,\tilde{c}+5} = \delta_{pc}$ with $c = \tilde{c} + 5$.

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.

E. Returns to reputation and ability by program-cohort

Our regression analysis in Section II is derived from a two-step estimation procedure. The first step equation (6) estimates conditional returns to reputation and ability separately for each program and cohort. The second step equation (7) relates these returns to the availability of the exit exam, captured in our treatment variable, δ_{pc}. Our benchmark specification (8) combines these two steps into a single regression.

To illustrate this procedure, Table B5 presents program-cohort specific returns from a regression similar to the first-step specification (6). Columns (A)-(C) display the 39 programs in our sample and the introduction year of the exit exam field we assigned to each program (see Table B1). Columns (F) and (G) present the conditional returns to reputation for each program and cohort, $\hat{r}_{pc}$, except we use only two cohort groups: students who graduated before the introduction of any exit exams (2003–2004) and those who graduated after the first field exams became available (2005–2009). Column (H) reports the difference between pre- and post-exam returns for each program. Columns (I)-(K) similarly show the program-cohort returns to ability, $\hat{a}_{pc}$, and their difference.

As shown in Table 2, most of our identification comes from a comparison of programs that received exit exams in the first year ("2004 programs") and programs that never received an exam during our period of analysis ("2009 programs"). We can thus illustrate our main results with a simple 2×2 difference in differences analysis using these two groups. The bottom rows of Table B5 show the average pre- and post-exam returns to reputation and ability for 2004 and 2009 programs.[54] The boxed numbers report the 2×2 difference in differences estimates. For example, the return to reputation declined from 13.8 to 9.8 percent in

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

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

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)
Exam Year	Program area	Program	N	Colleges	Return to reputation 2003–04	2005–09	Diff.	Return to ability 2003–04	2005–09	Diff.
2004	Agronomy	Medicina veterinaria	1,808	2	-0.09	-0.04	0.05	0.03	-0.02	-0.05
	Business	Administración	85,325	46	0.18	0.14	-0.05	0.04	0.03	-0.01
	Business	Contaduría pública	49,714	36	0.18	0.09	-0.09	0.03	0.03	0.00
	Business	Economía	25,879	21	0.21	0.11	-0.10	0.05	0.03	-0.02
	Education	Educación	40,195	21	0.10	0.05	-0.05	0.02	0.02	0.00
	Engineering	Ingeniería industrial y afines	41,309	25	0.23	0.15	-0.08	0.04	0.02	-0.02
	Engineering	Ingeniería de sistemas, telematica y afines	28,526	25	0.19	0.15	-0.04	0.05	0.04	-0.01
	Engineering	Ingeniería civil y afines	24,334	19	0.10	0.09	-0.01	0.02	0.01	-0.01
	Engineering	Ingeniería electrónica, telecom. y afines	15,657	14	0.19	0.13	-0.06	0.08	0.02	-0.06
	Engineering	Arquitectura y afines	11,701	12	0.07	0.05	-0.03	0.02	0.01	-0.01
	Engineering	Ingeniería mecánica y afines	9,659	9	0.27	0.19	-0.08	-0.02	0.01	0.03
	Engineering	Ingeniería ambiental, sanitaria y afines	7,251	8	0.14	0.06	-0.08	0.03	0.01	-0.02
	Engineering	Ingeniería agroindustrial, alimentos y afines	2,889	5	0.03	0.12	0.09	0.08	0.03	-0.05
	Engineering	Ingeniería química y afines	7,630	4	0.44	0.25	-0.19	0.02	0.06	0.04
	Engineering	Ingeniería eléctrica y afines	2,320	3	0.13	0.00	-0.13	0.01	0.03	0.01
	Engineering	Ingeniería agronómica, pecuaria y afines	1,559	3	0.30	0.26	-0.04	0.09	0.01	-0.07
	Health	Enfermería	27,824	19	0.06	0.08	0.02	0.01	0.01	0.00
	Health	Medicina	13,520	8	0.01	0.00	-0.01	0.03	0.01	-0.02
	Health	Terapias	12,211	8	0.05	0.03	-0.03	0.10	0.06	-0.04
	Health	Odontología	5,211	7	0.02	0.01	0.00	-0.03	0.00	0.02
	Health	Bacteriología	6,304	6	0.03	0.05	0.02	0.03	0.02	-0.01
	Health	Nutrición y dietética	3,635	3	-0.20	-0.22	-0.03	-0.03	-0.01	0.02
	Health	Optometría, otros prog. de ciencias de la salud	1,895	3	0.08	-0.01	-0.09	-0.01	0.00	0.02
	Social sciences	Psicología	35,506	24	0.11	0.07	-0. 04	0.03	0.02	-0.01
	Social sciences	Derecho y afines	39,608	21	0.12	0.10	-0.02	0.02	0.01	-0.01
	Social sciences	Comunicación social, periodismo y afines	19,523	16	0.18	0.13	-0.05	0.02	0.02	0.00
	Social sciences	Sociología, trabajo social y afines	7,442	7	0.08	0.05	-0.03	0.02	0.01	-0.01
2005	Natural sciences	Biología, microbiología y afines	7,418	5	0.04	0.17	0.13	0.01	-0.02	-0.02
2006	Health	Instrumentación quirúrgica	4,516	5	-0.22	0.03	0.26	0.02	0.02	0.01
2009	Engineering	Ingeniería administrativa y afines	3,936	5	0.16	0.12	-0.04	0.07	0.03	-0.04
	Engineering	Ingeniería de minas, metalurgia y afines	2,367	2	-0.01	-0.16	-0.15	0.13	0.02	-0.11
	Engineering	Otras ingenierías	1,558	2	0.11	0.13	0.02	-0.08	-0.02	0.05
	Fine arts	Diseño	12,641	7	0.02	0.04	0.02	0.04	0.01	-0.02
	Fine arts	Publicidad y afines	3,412	5	-0.01	0.00	0.02	0.03	0.02	-0.01
	Fine arts	Artes plásticas, visuales y afines	6,704	4	-0.19	-0.15	0.03	0.05	0.03	-0.03
	Social sciences	Ciencia política, relaciones internacionales	4,806	4	0.04	0.09	0.05	0.04	0.01	-0.03
	Social sciences	Lenguas modernas, literatura, ling. y afines	3,101	4	0.18	0.13	-0.05	0.10	0.03	-0.07
	Social sciences	Antropología, artes liberales	2,160	3	-0.22	0.04	0.26	0.04	-0.03	-0.07
	Social sciences	Geografía, historia	748	2	-1.84	21.05	22.89	0.25	0.00	-0.25
		2004 programs	528,435	94	0.138	0.098	-0.041	0.030	0.021	-0.009
		2009 programs	41,433	21	0.030	0.030	0.000	0.048	0.018	-0.030
		Difference			0.109	0.068	-0.041	-0.018	0.003	0.021

Note: Column (A) lists the introduction year of the exit exam field assigned to each of the 39 programs in our sample, which appear in column (C). Column (B) contains the program area of each program. Column (D) shows the number of earnings observations in our sample, and column (E) shows the number of colleges in our sample offering each program. See Table B1 and Appendices B.A and B.C for details.

Columns (F) and (G) report conditional returns to reputation for each program from specification (6) using only two cohort groups: 2003–2004 and 2005–2009. In other words, the returns to reputation coefficients are from a regression of log average daily earnings on interactions of reputation and Icfes with dummies for cells defined by programs and the 2003–2004 and 2005–2009 cohort groups. This regression includes an experience quadratic interacted with program dummies and dummies for program-cohort cells. Column (H) displays the difference between columns (F) and (G). Columns (I) and (J) report conditional returns to ability from the same specification, and column (K) displays their difference.

Averages at the bottom are weighted by each coefficient's inverse squared standard errors from this regression.

2004 programs, but was unchanged at 3.0 percent in 2009 programs. The difference in differences estimate is thus roughly -4 percent, similar to our benchmark coefficient in Table 3. The 2×2 estimate for the return to ability is 2.1 percent, which is also close to our benchmark result.

Table B5 also helps to explain the estimates in columns (E) and (F) of Table 3. These estimates restrict identification to programs with similar pre-exit exam returns to reputation and ability. Columns (F) and (I) in Table B5 show these pre-exam returns.[55] 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.

F. Exit exam effects on the returns to other characteristics

An alternative hypothesis for our main results is that the exit exams affected signaling on observable characteristics other than college reputation. To explore this hypothesis, in Table B6 we replicate our benchmark regression (column (A) in Table 3) replacing the reputation terms with other individual characteristics that may be at least partially observable to employers.

Column (A) replicates our benchmark results using college reputation. Columns (B)-(D) replace reputation with indicators for gender, mother's education, and family income, respectively. In each regression, the return to these other characteristics declines with the exit exam, but none of the effects are statistically significant. Further, the Icfes effects in all these regressions are also statistically insignificant. In column (E), we include terms for all characteristics jointly; only the Icfes and reputation effects are statistically significant.

Although we cannot rule out signaling effects on characteristics not included in our data, the results in Table B6 provide evidence that the strongest effects of the exit exams were on the returns to college reputation.

G. Balance tests

Section II.F discusses three balance tests that ask if the exit exam rollout was correlated with sorting into colleges or programs, or with the probability of formal employment. Table B7 shows the results from these balance tests. These estimates are from simple differences in differences regressions that include program dummies, cohort dummies, and our indicator for exposure to the exit exams, δ_{pc}. The dependent variable for each regression is listed in the column header.

In columns (A) and (B), the dependent variables are college reputation, R_s, and Icfes percentile, τ_i. If the field-specific introduction of the exit exam was

[55] In actuality, the pre-exit exam returns in Table B5 are estimated in a regression that also includes 2005–2009 graduates, while the pre-exit exam returns used for columns (E)-(F) of Table 3 are from a specification including only 2003–2004 cohorts. This has little effect on the returns displayed in Table B5.

Table B6—Exit exam effects on the returns to other characteristics

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)
Icfes × δ_{pc}	0.017	0.006	0.007	0.007	0.018
	(0.006)	(0.006)	(0.006)	(0.008)	(0.006)
Reputation × δ_{pc}	−0.041				−0.036
	(0.017)				(0.016)
Male × δ_{pc}		−0.021			−0.023
		(0.015)			(0.015)
College mother × δ_{pc}			−0.036		−0.026
			(0.035)		(0.036)
High income × δ_{pc}				−0.039	−0.031
				(0.030)	(0.038)
N	581,802	581,645	576,945	576,332	574,803
R^2	0.258	0.232	0.236	0.237	0.263
# programs	39	39	39	39	39
Mean return to ability	0.029	0.068	0.062	0.064	0.024
Mean return to reputation	0.133				0.125
Mean return to gender		0.038			0.038
Mean return to mother's ed			0.123		0.076
Mean return to income				0.115	0.066

Note: All regressions are identical to the benchmark specification in column (A) of Table 3, but they substitute the reputation terms in this regression with other characteristics. All columns report coefficients on the interactions of these characteristics with the treatment variable δ_{pc}. Regressions include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of Icfes and other characteristics with program and cohort dummies. Parentheses contain standard errors clustered at the program level.
College mother is an indicator for a student's mother having a technical college or university degree. High income is an indicator for a student's family income being greater than 300 percent of the minimum wage.
The mean returns at the bottom of the table are calculated using only 2003–2004 graduates.

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.[56] The mean of this variable is 66 percent, and the remaining 34 percent is a composite measure of unemployment, informal employment, non-participation in the labor market, and pursuit of further education. The estimate suggests that formal employment increased 1.7 percentage points more in programs with exit exam

[56] 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 B7—Balance tests

	(A)	(B)	(C)
	Dependent variable		
	Reputation	Icfes	Has formal earnings
Exposed to exit exam (δ_{pc})	0.026	0.070	0.017
	(0.051)	(0.078)	(0.016)
N	146,052	146,052	890,809
R^2	0.204	0.146	0.044
# programs	39	39	39
Dependent variable mean	7.44	7.84	0.66

Note: All columns report coefficients on the treatment variable δ_{pc}. Parentheses contain standard errors clustered at the program level.
The dependent variables in columns (A) and (B) are reputation and Icfes. The sample includes all students from Table 2. Each regression includes program dummies and cohort dummies.
The dependent variable in column (C) is an indicator for appearing in our earnings records at each year in 2008–2012. We include multiple observations per student for any level of potential labor market experience in 0–9 years. The sample includes all students from Table 2 plus graduates from the same programs and colleges who never appear in the earnings records. The regression includes program dummies, cohort dummies, and a quadratic in experience interacted with program dummies.

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.

H. Section III sample

In Section III, 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 B8 divide 2008–2009 graduates according to their post-college labor market paths. We choose these cohorts because our earnings records cover 2008–2012, which allows us to observe earnings in the year of graduation and the next three years.

Column (A) includes any student who enrolled in a specialization, masters, or doctorate program by 2011, the last year for which we have graduate education records. Columns (B)-(D) categorize those who did not enter graduate school by the number of years for which they have formal earnings in the first four years after graduation.[57] 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.

[57] We consider workers as having formal earnings if they have at least one monthly earnings observation in a given year.

Table B8—Transition from college to the labor market

2008–2009 college graduates

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

Note: The sample includes 2008–2009 graduates from the sample for Figure 1. We choose the 2008–2009 graduation cohorts so that we observe earnings for the first four years after graduation (2008–2011 for 2008 graduates, and 2009–2012 for 2009 graduates).
Column (A) includes any student who enrolled in a specialization, masters, or doctorate program in 2007–2011, the years for which we have graduate education records from the Ministry of Education. Column (B) contains non-graduate school students who never appear in our earnings records in the first four years after graduation. Column (C) contains non-graduate school students who appear in the earnings records in some but not all of the first four years. Column (D) contains non-graduate school students who appear in our earnings records in all four years.
Parentheses contain standard deviations. College educated mother is a dummy equal to one if a student's mother has a college/postgraduate degree.

Column (A) 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.[58]

Our sample for Section III 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.

I. *Unconditional return to ability*

This section presents results related to the Proposition 2 predictions for the unconditional return to ability (Icfes).[59]

[58] F-tests for each characteristic strongly reject the hypothesis of joint equality across the four columns.
[59] We note that the Icfes percentiles we use in Section III are conceptually similar to those in Section II, but they are based on different data sources. In Section III, we compute Icfes percentiles using data

Table B9—Returns to ability and experience interactions

	Dependent variable: log average daily earnings	
	(A)	(B)
Icfes	0.045	0.027
	(0.006)	(0.004)
Icfes × t	0.009	0.003
	(0.001)	(0.001)
N	83,492	83,492
R^2	0.163	0.297
# colleges	130	130
Extra controls		Y

Note: The sample includes students in column (D) of Appendix Table B8 and earnings in the four years after graduation. Column (A) estimates equation (10) excluding reputation terms. In addition to the reported variables, the regression includes dummies for cohort-experience cells.
Column (B) adds the following controls to column (A): age at graduation, a gender dummy, dummies for eight mother's education categories, dummies for missing age and mother's education values, college program dummies, and dummies for college municipalities. Each control is interacted with a quadratic in experience.
Parentheses contain standard errors clustered at the college level.

Column (A) of Table B9 estimates (10) including Icfes terms but not reputation terms, such that the coefficients represent the unconditional returns to ability, a^u (equation (5), Section I). The coefficient on Icfes shows that a ten percentile increase in the student's score is associated with a five percent increase in daily earnings in the year of graduation ($a_0^u \approx 0.05$). The standard deviation of Icfes percentiles is about twice that of reputation, and hence scaled by this measure the unconditional returns to reputation and ability are of a similar magnitude.

Proposition 2 states the coefficient on Icfes should increase with experience, i.e., it predicts a positive coefficient on the interaction of Icfes and experience. This follows from the assumption that employers do not fully observe Icfes scores, and thus the correlation of wages and Icfes increases as workers reveal their skill through their output. Column (A) is consistent with this prediction. The point estimate on the Icfes-experience interaction implies that the return to ability grows by roughly 60 percent in the first four years after graduation.

Column (B) adds controls for graduates' gender, age, socioeconomic status, college program, and regional market. All controls are interacted with a quadratic in potential experience to allow earnings trajectories to vary with each characteristic. The coefficient on the Icfes-experience interaction decreases slightly, but it is still highly significant.

from the Colombian Institute for Educational Evaluation (see the notes to Figure 1). This yields a relatively continuous variable. In Section II, 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.

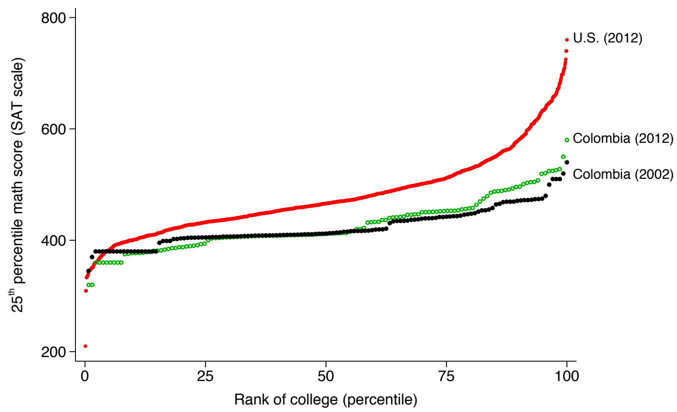

Figure B9. Ability sorting in Colombia and the U.S.

Note: The y-axis shows the 25^{th} percentile math scores for entering students at U.S. and Colombian colleges. The x-axis depicts unweighted percentile ranks using these 25^{th} percentile math scores. U.S. SAT math percentiles are from the Integrated Postsecondary Education Data System. We include 1,271 four-year degree-granting public and private not-for-profit colleges with ten or more 2012 first-time degree/certificate-seeking undergraduates. Colombian colleges are the same as in Figure 1 (except three have no 2012 enrollees). We include students who enrolled in either 2002 or 2012 and took the Icfes no more than two years before enrolling. We calculate Icfes math percentiles relative to the enrollment cohorts and convert them to an SAT scale using the distribution of math scores for 2011 U.S. college-bound seniors, available in January 2015 at http://media.collegeboard.com/digitalServices/pdf/SAT-Mathemathics_Percentile_Ranks_2011.pdf. We jitter interior 25^{th} percentile math scores slightly to smooth out discrete jumps in SAT scores.

The increasing return to ability is similar to the Farber and Gibbons (1996) and Altonji and Pierret (2001) findings using AFQT scores as an unobserved characteristic. However, it is in contrast with findings in Arcidiacono, Bayer and Hizmo (2010), who also study AFQT scores but make a distinction between graduates who enter the labor market after high school and those who do so after college. For college graduates, they show that AFQT is strongly related to wages in the year of graduation, and this relationship changes little over the next ten years. Their conclusion is that AFQT revelation is complete for college graduates, and they suggest that this revelation occurs through college identity.

The difference in findings may be explained by the fact that sorting by ability in Colombia—although increasing—appears to be less extensive than in the U.S. Specifically, if the U.S. experience is indicative, one might expect sorting by ability to increase in Colombia as reductions in the cost of transport and information gradually move regional college markets away from relative autarky (Bound et al., 2009; Hoxby, 2009).

Figure B9 illustrates these dynamics in Colombia and its current standing relative to the U.S. We first plot the 25^{th} percentile Icfes math scores in the 2002 and 2012 entering cohorts at each college, with schools ranked on the x-axis according to this 25^{th} percentile.[60] 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 similar decline.

Despite these dynamics, by this measure Colombia's college market features substantially less sorting than that in the U.S. Figure B9 also shows the 25^{th} percentile SAT math scores for the 2012 entering cohort at U.S. four-year degree-granting public and private not-for-profit colleges. Comparing Icfes and SAT scores requires strong assumptions, as the tests may capture different characteristics, but 25^{th} percentile math scores increase much more rapidly in the U.S. While both countries have colleges with 25^{th} percentile scores below 400 SAT points, the top-ranked U.S. colleges are above 700, and no Colombian college surpasses 600.[61]

A plausible explanation for the positive coefficient on the interaction of Icfes and experience in Table B9 is thus incomplete sorting by ability across Colombian colleges. The more substantial sorting by ability across U.S. colleges may result in a more complete reflection of AFQT in wages upon graduation.[62]

J. *Return to years of schooling in Colombia*

Our main result from Section III is that the return to college reputation in Colombia increases with experience. This differs from the standard U.S. result that the return to years of schooling does not change with experience. This subsection shows that this benchmark years of schooling finding also holds in Colombia, as previewed in Panel A of Figure 4.

For this we use cross-sectional data from the 2008–2012 monthly waves of the Colombia Integrated Household Survey (*Gran Encuesta Integrada de Hogares*). This survey measures workers' hourly wages and years of schooling, which range from 0–20 years. We calculate each worker's potential experience, t, as $t = \min(\text{age} - \text{years of schooling} - 6, \text{age} - 17)$, and include workers with experience levels 0–39.[63]

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

[61] If we convert Icfes scores to an SAT scale using the entire population of Icfes takers—instead of only those who entered college—the dots describing Colombia in Figure B9 shift up and become somewhat steeper, but they still exhibit a flatter slope than exists for U.S. colleges. This renormalization, however, overstates the amount of sorting in Colombia relative to the U.S. because Icfes test takers are less likely to enroll in college than SAT test takers. Using only college enrollees to make this conversion is more appropriate because the distribution of SAT scores we use is for U.S. college-bound seniors.

[62] If we estimate Table B9 with Icfes scores normalized to mean zero and standard deviation one—as Arcidiacono, Bayer and Hizmo (2010) do with AFQT—the period zero coefficient on Icfes is approximately one half of their AFQT coefficient. Although the two tests may measure different individual characteristics, the relative magnitudes are also consistent with partial revelation of the ability of college graduates in Colombia.

[63] We note that this definition of potential experience differs from the one we use elsewhere in the

Table B10—Return to years of schooling and experience interaction

2008–2012 cross-sectional household survey

	(A)	(B)	(C)	(D)
	Dependent variable: Log hourly wage		Dependent variable: Log weekly earnings	
	0–39 years experience	0–9 years experience	0–39 years experience	0–9 years experience
Years of schooling	0.1224	0.1239	0.1150	0.1192
	(0.0008)	(0.0018)	(0.0009)	(0.0021)
Years of schooling × t	−0.0002	−0.0001	−0.0002	−0.0006
	(0.0000)	(0.0003)	(0.0000)	(0.0003)
N	660,573	217,523	660,573	217,523
R^2	0.407	0.352	0.351	0.308

Note: 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.

Table B10 shows how the return to years of schooling in Colombia changes with experience. The use of cross-sectional data differentiates Table B10 from the panel data results in Farber and Gibbons (1996), Altonji and Pierret (2001), and Table 7 of this paper, but it is similar to the original Mincerian regressions that rely on U.S. survey data (e.g., Lemieux, 2006).

Column (A) displays the coefficients from a regression of log hourly wages on years of schooling and its interaction with experience.[64] The results suggest that an additional year of education is associated with a 12 percent increase in initial wages, and that this gap remains roughly constant as workers gain experience. The coefficient on the interaction term is statistically significant due to the large sample size, but it is close to zero. For example, after ten years the return to schooling decreases by only 0.002 log points, or less than two percent of the initial return.

Column (B) of Table B10 restricts the sample to workers with 0–9 years of potential experience, with negligible impact on the results. This matches the experience levels we can observe using our administrative data on Colombian college graduates, as depicted in Panel B of Figure 4.

Columns (C)-(D) of Table B10 replicate columns (A)-(B) with log weekly earnings (rather than log hourly wage) as the dependent variable. This is motivated

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).

[64] Regressions in Table B10 also include controls for experience and survey date.

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.

K. Robustness of increasing return to reputation

Table B11 documents the robustness of our main result from Section III: the return to reputation—even conditional on Icfes scores—increases with experience (column (B) of Table 7). As a benchmark, we reproduce this result in column (A) of this table. The sample for this regression includes students from column (D) of Table B8. We regress log average daily earnings on dummies for cohort-experience cells, reputation, Icfes, and the interactions of both variables with experience. The point estimate on the reputation-experience interaction suggests that the effect of a one unit increase in reputation on earnings grows by about 1.2 percentage points each year.

Columns (B)-(D) test the sensitivity of this result to the addition of controls. Column (B) adds controls for gender, age at graduation, and socioeconomic status as measured by mother's education. We interact all variables with a quadratic in experience so that controls can affect both the intercept and the slope of graduates' earnings profiles. The addition of these controls for personal characteristics lowers the coefficient on the interaction of reputation and experience slightly, though it is still significant and roughly the same magnitude in proportion to the period-zero return to reputation.

Column (C) includes all controls from column (B) and adds two characteristics of graduates' colleges. First, we add dummies for college programs (see column (C) of Table B1) and their interaction with a quadratic in experience. These dummies are important if graduates from different programs enter occupations that vary in their potential for wage growth. Second, we add dummies for college municipalities and the interactions of these dummies with an experience quadratic. Location controls may matter if earnings paths differ across regional markets. Our estimates in column (C) are thus identified off of variation in college reputation for students in the same programs and cities. The magnitude of the reputation-experience coefficients falls again, but it is still significant and is slightly larger in relation to the initial return to reputation.

In addition to the controls in column (C), column (D) adds each graduate's log earnings in the year of graduation. The inclusion of experience-zero earnings is in the spirit of Farber and Gibbons (1996), who use initial wages to control for other worker characteristics observable to employers but not to the econometrician. We additionally interact initial earnings with a quadratic in experience to control

Table B11—Alternate specifications for return to reputation and experience interaction

	Dependent variable: log average daily earnings						
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
		Additional controls & experience interactions			Degrees of labor market attachment		
	Benchmark estimates	Gender, age, & SES	Program & municipality	Initial earnings	Actual experience	Full-time employment	No prior employment
Reputation	0.079	0.072	0.055	0.007	0.066	0.086	0.067
	(0.017)	(0.015)	(0.016)	(0.001)	(0.017)	(0.017)	(0.020)
Reputation × t	0.012	0.010	0.008	0.016	0.015	0.015	0.019
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)
Icfes	0.024	0.022	0.017	0.001	0.018	0.026	0.027
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.007)
Icfes × t	0.006	0.004	0.002	0.004	0.006	0.007	0.007
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
N	83,492	83,492	83,492	83,492	83,492	39,596	7,168
R^2	0.190	0.203	0.313	0.627	0.242	0.230	0.230
# colleges	130	130	130	130	130	130	113
Personal traits × $f(t)$		Y	Y	Y			
College traits × $f(t)$			Y	Y			
Initial earnings × $f(t)$				Y			
Definition of t	Potential	Potential	Potential	Potential	Actual	Potential	Potential
Full-time restriction						Y	Y
Prior work restriction							Y

Note: All columns report coefficients on reputation, Icfes, and their interactions with experience. The sample includes the 2008–2009 graduates from column (D) of Appendix Table B8 and earnings within four years after graduation. All regressions include dummies for cohort-experience cells. Parentheses contain standard errors clustered at the college level.

Column (A) is identical to column (B) in Table 7. Columns (B)-(D) layer in additional controls, and every variable we add is interacted with a quadratic in experience. Column (B) adds a gender dummy, age at graduation, dummies for eight mother's education categories, and dummies for missing age and mother's education values. Column (C) includes all controls in column (B) plus program dummies and dummies for college municipalities. Column (D) includes all controls in column (C) plus log average daily earnings at experience zero.

Columns (E) is identical to column (A), but all experience terms are defined using actual experience—the cumulative number of months with earnings since graduation—rather than potential experience. Column (F) is identical to column (A), but we include only graduates who have earnings in every month starting in the year after graduation. Column (G) includes only those students in column (F) who do not appear in our earnings records in the year prior to graduation. This column includes only 2009 graduates, for whom we can observe pre-graduation employment.

for variation in earnings trajectories across jobs with different starting wages. The controls for initial earnings mechanically reduce the period-zero reputation and Icfes coefficients, but the coefficient on the interaction of reputation and experience doubles in magnitude relative to column (C).

In columns (E)-(G), we remove the controls from columns (B)-(D) and instead test the sensitivity of our result to the degree of graduates' labor market attachment. As discussed, the sample for Table B11 includes only students who are employed in each of the first four years after graduation, but graduates may still differ in the number of months they are employed in each year. In all previous specifications, we measure labor market experience using potential experience, defined as calendar year minus graduation year. Column (E) of Table B11 is identical to column (A), but we replace all experience terms with *actual* experience, defined as the number of months of employment since graduation.[65] This alternate measure may be important if graduates from high reputation colleges are more likely to find stable employment, but the results in column (E) are similar to our benchmark estimates.

Column (F) is identical to column (A), but we restrict the sample to include only students who have full-time employment after graduation. In column (A) we require that each student have at least one monthly earnings observation in each of the first four years after graduation. In column (F), students must have an earnings observation in *every* month beginning in the year after graduation. This requirement reduces the sample size by more than 50 percent but has little effect on the reputation-experience coefficient.

Column (G) makes a further restriction to the sample from column (F). In this column we also require that graduates were not employed in the year *before* graduation. This restriction may be important if graduates from top colleges are less likely to work while in school, and if prior employment affects future wage growth. Since our earnings records begin in 2008, we can only observe pre-graduation employment for 2009 graduates. Thus, column (G) includes only 2009 graduates who have no earnings in 2008. This restriction leads to a small sample in column (G), but if anything, the coefficient on the interaction of reputation and experience is larger in this population.

In sum, Table B11 suggests that the increasing conditional return to reputation is not driven by variation in earnings paths across individual characteristics, college programs, regional markets, or levels of initial earnings. Furthermore, this result does not appear to stem from variation across colleges in labor market attachment.

[65] Papers in the employer learning literature use different measures of experience and potential experience. Farber and Gibbons (1996) use experience based on actual employment duration, while Altonji and Pierret (2001) principally use potential experience based on age and years of schooling. Potential experience based on graduation year is most logical for our study of college reputation and is consistent with the primary measure used by Arcidiacono, Bayer and Hizmo (2010).

L. College-program level reputation

This section provides details on the robustness of our main results using a college-program level definition of reputation rather than a college level definition. Table B12 replicates the main exit exam results from Table 3 with reputation defined as college-program mean Icfes. The results closely mirror our main findings in sign and magnitude, although the standard errors are typically larger. This is likely due to the fact that the college-program reputations are calculated from smaller samples. In general this does not alter the pattern of statistical significance relative to Table 3, with the exception of statistically insignificant reputation effects in columns (B) and (D).

Table B13 replicates the results on earnings growth from Table 7 in Section III with college-program level reputation. The main findings are unaltered by this modification. In particular, the coefficient on the reputation-experience interaction is positive and highly significant in all specifications.

Table B12—Exit exam effects with college-program level reputation

	Dependent variable: log average daily earnings					
	(A)	(B)	(C)	(D)	(E)	(F)
		Experience & cohort controls		Restriction to similar programs		
	Benchmark specification	Within experience	Linear trends	S. sciences & engineering	Within $\hat{r}_p$ quartiles	Within $\hat{a}_p$ quartiles
Reputation × δ_{pc}	−0.038	−0.021	−0.015	−0.052	−0.026	−0.054
	(0.022)	(0.019)	(0.028)	(0.031)	(0.012)	(0.023)
Icfes × δ_{pc}	0.019	0.017	0.010	0.043	0.014	0.009
	(0.007)	(0.009)	(0.009)	(0.012)	(0.007)	(0.004)
N	581,802	267,924	267,924	273,590	581,802	581,802
R^2	0.258	0.225	0.225	0.263	0.259	0.259
# programs	39	39	39	22	39	39
Experience levels	0–9	4–7	4–7	0–9	0–9	0–9

Note: This table is identical to Table 3 in Section II, but it uses reputation defined as mean Icfes at the college-program level rather than at the college level. All columns report coefficients on the interactions of reputation and Icfes with the treatment variable δ_{pc}. Regressions in columns (A) and (C)-(F) include a quadratic in experience interacted with program dummies, dummies for program-cohort cells, and interactions of both reputation and Icfes with program and cohort dummies. Column (B) includes dummies for program-cohort-experience cells and interactions of both reputation and Icfes with program-experience and cohort-experience dummies. The sample for each regression is restricted to the experience levels listed in the bottom row. Parentheses contain standard errors clustered at the program level.

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

Table B13—Experience interactions with college-program level reputation

Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)
Reputation	0.064	0.039	0.044	0.024
	(0.015)	(0.014)	(0.014)	(0.013)
Reputation × t	0.012	0.012	0.008	0.004
	(0.002)	(0.002)	(0.002)	(0.001)
Icfes		0.045	0.034	0.023
		(0.005)	(0.004)	(0.004)
Icfes × t			0.007	0.002
			(0.001)	(0.001)
N	83,492	83,492	83,492	83,492
R^2	0.156	0.178	0.179	0.301
# colleges	130	130	130	130
Extra controls				Y

Note: This table is identical to Table 7 in Section III, but it uses reputation defined as mean Icfes at the college-program level rather than at the college level. The dependent variable is log average daily earnings. The sample includes students in column (D) of Appendix Table B8 and earnings in the four years after graduation. Columns (A)-(C) estimate equation (10) excluding and including Icfes terms. In addition to the reported variables, both regressions include dummies for cohort-experience cells.
Column (D) adds the following controls to column (C): age at graduation, a gender dummy, dummies for eight mother's education categories, dummies for missing age and mother's education values, college program dummies, and dummies for college municipalities. Each control is interacted with a quadratic in experience.
Parentheses contain standard errors clustered at the college level.