THE BIG SORT: COLLEGE REPUTATION AND LABOR MARKET OUTCOMES

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

AUGUST 27, 2015

Abstract. Firm reputation—defined as the quality buyers expect from a seller—can address transactional problems in markets with imperfect information. This paper studies the role college reputation plays in matching students to schools and transmitting their ability to employers. We incorporate college reputation into a model of wage formation and propose a measure of reputation that yields a clean test of signaling. We combine administrative data and a natural experiment from Colombia to show that reputation indeed signals ability. Finally, we show that graduates' college reputations are positively correlated with their earnings growth, suggesting that reputation matters beyond signaling.

For useful comments we thank Joseph Altonji, Costas Meghir, Michael Mueller-Smith, Phil Oreopoulos, Kiki Pop-Eleches, and Russell Weinstein. For invaluable help with the data we are grateful to Julian Mariño and Adriana Molina at the Colombian Institute for Educational Evaluation (ICFES), Luz Emilse Rincón at the Ministry of Social Protection, and Luis Omar Herrera at the Ministry of Education. All errors are ours.
*Columbia University and NBER, +Columbia University, #University of Southern California.

1. Introduction

Firm reputation—defined as the quality buyers can expect from a seller—can help address transactional problems in markets with imperfect information (Akerlof, 1970; Shapiro, 1983).

Informational issues are particularly salient in college markets, as college quality is relevant to two types of buyers. The first is the student. For her the quality of a college depends on the experience it provides her, and on the effect is has on her career. Ideally, she would want to know this ex-ante for all colleges in her choice set. Since this is not feasible, she must choose based on expectations that are shaped by colleges' reputations.

This paper examines college reputation from the perspective of a second type of buyer: employers. We ask if employers use college identity to infer the ability of individual graduates. This builds on Tirole's (1996) observation that individuals in a given group impose reputational externalities on each other, although as in Coate and Loury (1993), our focus is on the transmission of information about personal characteristics rather than past behavior. More directly, we explore empirical implications from MacLeod and Urquiola (2015), who show that signaling mechanisms can lead students to prefer colleges with high ability peers.

We incorporate college reputation into a model of wage formation to study the role it plays in matching students to colleges and to careers. We define a college's reputation to be the mean admission score of its graduates. This measure captures only one dimension of expected employee quality: ability as measured by admission exams. But we show that it yields a clean test of signaling if one's data also include individual admission scores.

We then bring together two ingredients necessary to implement such a test: data and identification. First, we link unique administrative information on college graduates in the country of Colombia. We observe each student's score on a national admission exam, her college of graduation, and her labor market outcomes for several years. Second, we exploit a natural experiment from Colombia's introduction of national college exit exams, which provided the market with a new, individual-specific measure of skill. We ask how this change in employers' information set affects the correlations of earnings with the group-level measure of ability—college reputation—and the individual measure—admission scores.

Our identification comes from the staggered rollout of exit exams across 55 fields like accounting, dentistry, economics, and law. This rollout allows us to implement an approach analogous to Card and Krueger (1992), who analyze how time-varying state policies (e.g., class size levels) can affect a slope—the relation between years of schooling and wages. In our case the question is how time-varying college major characteristics (e.g., the existence of an exit exam in a related field) affects two slopes—the earnings return to reputation and the earnings return to admission scores. If employers use college reputation to infer individual ability, then a new signal of skill should reduce the return to reputation and increase the return to admission scores. This is exactly what we find.

In addition, we present suggestive evidence that the exit exams improved overall employer-employee match quality as measured by average earnings. We also find that they prompted student behavioral responses in the form of delayed graduation and preference for colleges and programs with better exit exam performance.

These results provide evidence that college identity signals ability. We then ask whether college reputation relates to earnings *exclusively* through signaling. We show that if reputation is purely a signal of ability as measured by admission scores, then in a competitive labor market its effect on wages should not change as workers gain experience. Moreover, since we have defined reputation as the mean admission score at a college, it contains no information on "admission exam" ability conditional on individual scores. This implies that as individuals reveal their ability through their job performance, college reputation should have a declining relationship with wages conditional on admission scores.

Our results reject these predictions. Even after controlling for admission scores, graduates' starting earnings and earnings *growth* are both positively correlated with the reputation of their college. These results are consistent with the hypothesis that colleges add to skill, and that their value added varies systematically with their reputation. Although we cannot establish that this is a causal link, these correlations matter because they are observable; if students observe that individuals from better schools get careers with higher earnings trajectories, this may lead them to prefer more reputable schools.

Our paper relates to four distinct literatures: reputational markets, college choice, the impact of selective schools, and signaling/wage dynamics.

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

College choice. Hoxby (1997, 2009) finds that stratification by ability has increased significantly among colleges in the U.S. Thorough sorting may account for the fact that Arcidiacono et al. (2010) find that college identity in the U.S. seems to fully reveal Armed Forces Qualification Test (AFQT) scores. In contrast, we find that college identity only partially reveals admission test scores in Colombia. This may reflect that stratification there, although increasing, is not as thorough as in the U.S. In addition, Hoxby and Avery (2013) show that even controlling for ability, individuals from disadvantaged backgrounds are less likely to apply to reputable colleges.[1] This suggests that college preferences, and hence reputations, are endogenous. Our results are consistent with this hypothesis in that the introduction of exit exams altered the labor market implications of college reputation, and the preferences of college applicants. The latter effect is also relevant to work on matching in college and other markets (e.g., Roth and Sotomayor, 1989; He, 2014).[2] These useful models assume that students have clear preferences over the colleges they wish to attend; instability in these preferences might be an area to explore.

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

Signaling and wage dynamics. Spence (1973) noted that if there is signaling, a college wage premium can exist even if college has no value added. We have explored analogous issues when the question is *which* college students attend rather than *whether* they attend. This raises mechanisms that differ from those in Spence (1973). In his framework the key driver is that schooling is costly, but less so for individuals of high ability; in equilibrium, therefore, only the most able go to college. In contrast everyone in our data is a college graduate. In Spence (1973) there is no rationing; implicitly a single school sets a difficulty level and accepts anyone who wishes to attend. In our setting there are many selective colleges.

[1] More generally, Avery et al. (2013) provide a way to estimate the revealed preferences of students over colleges. Their results are generally consistent with a role for "brand name" reputations.

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

Jovanovic (1979) introduced the assumption that wages in a competitive market reflect all available information regarding worker skill. Farber and Gibbons (1996) and Altonji and Pierret (2001) find that as workers gain experience, observable characteristics like years of schooling become less correlated with wages in regressions that include unobserved measures of ability.[3] This suggests that schooling signals ability, while other factors correlated with schooling have a deterministic effect on wages. We find that even controlling for individual admission scores, college reputation is positively correlated with graduates' initial earnings and earnings *growth*—a starkly different pattern. Our findings suggest that the sorting that takes place by educational quality differs from that which takes place by quantity.

Finally, there is also work considering program of study. Grogger and Eide (1995) and Arcidiacono (2004) find that there is considerable variation in returns by major in the U.S. Altonji et al. (2014) update this work, finding that choice of major, as opposed to individual ability, continues to have a significant effect upon earnings. In this paper we do not focus on program of study (although our regressions control for it).

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

2. College reputation, signaling, and wages

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

2.1. **Ability, admission scores, and college reputation.**

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

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

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

[3] Lange (2007) finds that errors regarding worker skill decline markedly after a few years of employment, although Kahn and Lange (2014) find greater persistence.

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

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

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

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

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

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

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

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

We suppose that employers' information set, I_{it}, includes college reputation, R_{s_i}.[4] While employers likely care about individuals' pre-college ability as captured by R_{s_i}, they also care about other attributes related to graduates' post-college skill. We therefore define a college's *labor market reputation* as the expected skill of its graduates: $\mathscr{R}_s = E\{\theta_i | i \in s\}$. It follows that $\theta_{i\in s} \sim N(\mathscr{R}_{s_i}, \frac{1}{\rho^{\mathscr{R}}})$, where $\rho^{\mathscr{R}} = \frac{1}{\sigma^2_{\mathscr{R}}}$ denotes the precision of $\mathscr{R}_s$.[5]

[4] 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.

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

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

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

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

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

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

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

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

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

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

where ϵ_{it} includes human capital growth and other fluctuations in worker output. These are observed *after* setting wages in each period t (where $t = 0$ is the year of graduation). Let $\bar{y}_{it} = \frac{1}{t+1}\sum_{k=0}^{t} y_{ik}$ denote mean worker output and let $\rho^{\bar{y}}$ be the time-invariant precision of y_{it}.[6]

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

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

[6] The assumption that the precision of y_{it} is time stationary also follows Farber and Gibbons (1996).

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

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

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

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

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

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

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

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

where $\beta_t^r = r_t^{exit} - r_t$ and $\beta_t^a = a_t^{exit} - a_t$. Appendix A.4 shows that $\beta_t^r < 0$ and $\beta_t^a > 0$. Thus we have:

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

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

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

Our definition also helps distinguish a signaling channel from competing hypotheses. For example, in our context college-mean exit exam scores were made available, which may have altered the labor market's perception of college reputation. This could explain a declining importance of our measure, R_s, but does not explain the shift in weight from R_s to individual admission scores. The exit exam may also have prompted institutional responses such as changes in curricula. This would affect skill formation while at college, included in v_s; it would not affect pre-college ability, α_i, the focus of our analysis.

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

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

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

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

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

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

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

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

Thus Proposition 2 allows us to test whether the return to reputation arises solely because college identity signals ability as measured by admission scores. This is akin to the classic Spence hypothesis in the context of educational quality rather than educational quantity. Rejection by the data would suggest that other college membership attributes lead reputation to be correlated with wage growth. We examine these hypotheses in Section 4.

3. The college exit exam

This section tests Proposition 1. We first discuss institutional background and our measure of reputation. We then turn to the exit exam, sample, empirical specifications, and results.

3.1. **Background and data sources.** Colombia's higher education system consists of public and private institutions that award various types of degrees. In this paper, we refer to "colleges" as institutions that award the equivalent of U.S. bachelor's degrees after four or five years of study. Colombia also has institutions that specialize in two or three year degrees. We set these aside to focus on institutional identity within a single schooling level.[7]

To apply to college, students are required to take a standardized exam, the Icfes.[8] The Icfes is generally analogous to the SAT, but it is taken by the vast majority of high school

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

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

seniors regardless of whether they intend to apply to college.[9] The Icfes plays a major role in college admissions: many schools extend admission offers based solely on students' Icfes performance; others consider additional factors, and a handful administer their own exams.

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

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

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

(3) The Ministry of Social Protection provided monthly earnings records for formal sector workers during 2008–2012. These come from data on contributions to pension and health insurance funds. We calculate average daily earnings by dividing base monthly earnings for pension contributions by the number of formal employment days in each month and averaging across months.[10] This agency also provided four-digit economic activity codes for the first job in which a worker appears in their records.

3.2. **Ability and college reputation.** We define two measures of ability that correspond to those in the theory (Section 2). The first is student i's score on the Icfes admission exam, which we denote by τ_i. Throughout, we express Icfes scores as percentiles relative to all high school seniors who took the exam in the same year. The second is the reputation of a college s, denoted by R_s, defined as the mean Icfes score of its graduates.[11] To avoid capturing any effects from the exit exam rollout on reputation, we calculate R_s using graduates who took the Icfes exam in 2000–2003.

Icfes and reputation are divided by ten so that both measures range from 0–10 and one unit is ten percentile points. One unit of reputation is about one standard deviation in this measure, and it is roughly sufficient to move from either the 75^{th} to the 100^{th} percentile, or

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

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

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

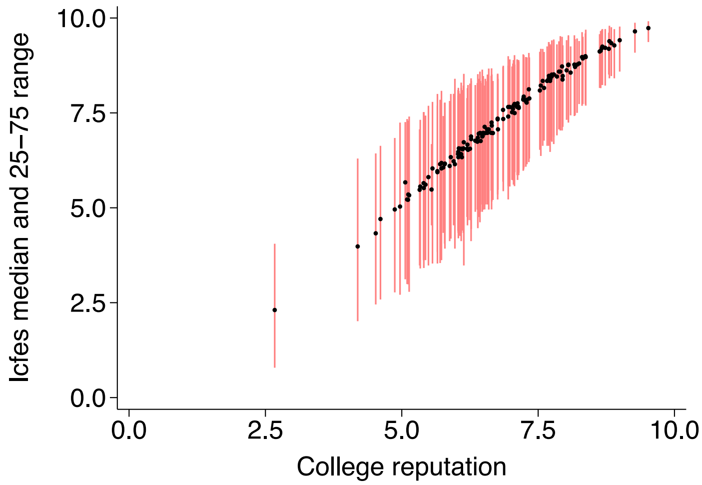


FIGURE 1. College reputation and individual ability

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

from the 50^{th} to the 75^{th}. Anecdotally, a student applying to a very top college might also apply to one with one point lower in reputation as a "safety school."

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

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

Rather than focus on its accountability dimension, we analyze the exit exam as potentially affecting students' capacity to signal their skill. This is consistent with anecdotal evidence that many students list exit exam scores on their CVs or on online profiles.[12] The exit exam may also affect faculty recommendations or students' search behavior after learning their position in the national distribution of exam takers.

3.4. **Identification.** To identify the effects of this new signal of skill, we exploit the gradual rollout of the exam fields in an "intent to treat" spirit. Exams were introduced in 55 fields between 2004 and 2007. The initial fields were those related to popular majors such as economics and industrial engineering; fields corresponding to less common degrees were introduced later (Appendix B.1 lists all fields and their introduction year). During this time the exams were not required, although they were taken by the majority of students in related majors. In 2009, the exit exam became mandatory for graduation, and a "generic competency" exam was made available for majors without a corresponding field.

Although the exit exams were field-specific, during the period we study there was no formal system assigning college majors to exam fields. This match is necessary to determine which majors were treated. We therefore perform this assignment ourselves using the Ministry of Education's 54 major groups, which we label programs.[13] We assign each of the 54 programs to one of the 55 exam fields if the program name appears in the name of the field exam. We assign programs without matching names to the generic competency exam introduced in 2009. Appendices B.1 and B.2 describe this matching procedure and show that our main results are robust to several alternative matching methods.

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

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

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

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

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

Notes: This table displays the match of college programs to the exit exam field and generic exam years. Programs are the Ministry of Education's 54 core knowledge groups, which are further categorized into the listed eight program "areas." Appendix B.1 lists the exam fields and details how we match them to programs.

We define a binary treatment variable δ_{pc}, which equals one if students in program p and graduation cohort c had an available exit exam in the matched field. Because students typically take the exam one year before graduating, the first treated cohort is that which graduated one year after the introduction of the field assigned to its program.[14] For example, $\delta_{pc} = 1$ for psychology students who graduated in 2005 or later because the psychology field exam was introduced in 2004. $\delta_{pc} = 0$ for all anthropology students who graduated before 2010 because the testing agency did not produce a related exam field.

Figure 2 shows that the introduction of exit exam fields led to sharp increases in the fraction of students taking the test. For example, the test taking rate in 2004 programs jumped from 10 to 55 percent with the 2005 cohort, the first we define as treated for this program group. Students in 2009 programs rarely took the exam until the cohort following the exit exam mandate in 2009.[15]

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

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

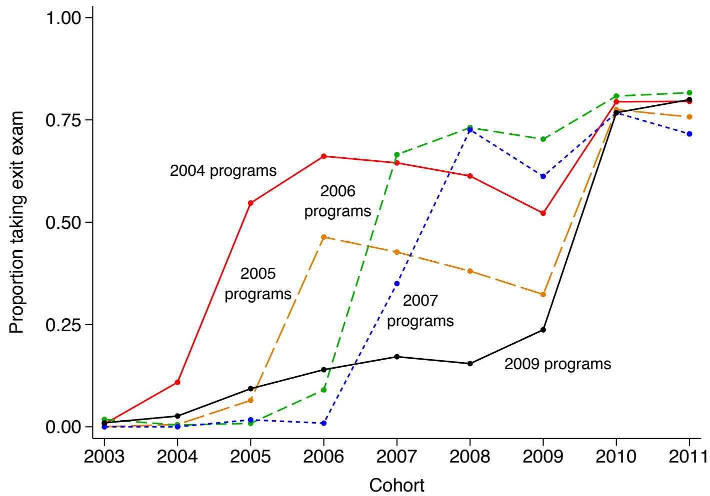


FIGURE 2. Proportion of students taking exit exam by program group

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

To summarize, we define a treatment indicator, δ_{pc}, at the program-cohort rather than at the individual level. Thus we analyze the introduction of the exams in an "intent to treat" spirit. This reflects that beyond the fact that students were not required to take exit exams during the period we study, they had no obligation to disclose their performance if they did (although not doing so might in itself convey information). Thus, while we can assert that the introduction of the exam into a student's field potentially affected the information available in that individual's labor market, we do not know precisely how it affected what firms observed about her.[16]

3.5. **Sample.** We analyze the effects of the exit exam using the 2003–2009 graduation cohorts. With these we can focus cleanly on the period in which signals of skill were introduced into a subset of fields.[17] Table 2 presents summary statistics separately for program groups defined by the year each program received its assigned exit exam field. Approximately 90 percent of students graduate from programs that received an exam field in 2004; most of the remaining graduates had no corresponding field until the 2009 generic exam.

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

[16] The potential endogeneity of exam taking also explains why we do not use the exit exam scores in our main analysis, either to define reputation or as a measure of graduates' skill.
[17] This is no longer clearly the case after the 2009 cohort due to several structural changes in the exit exams.

Table 2. Summary statistics for exit exam sample

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

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

Our sample includes 39 programs offered at 94 colleges. These numbers are smaller than the total number of programs defined by the Ministry of Education (54) and the number of colleges in our records (136). We exclude programs and colleges that have too few observations to precisely estimate a return to reputation among graduates from the same program—a necessity for our empirical specification below. Appendices B.3 and B.4 provide details on the sample selection and show that our main results are robust to the key restrictions.

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

The last two rows in Table 2 report the returns to reputation and ability (Icfes) within each program group. These are analogous to the r and a coefficients from equation (2) in Section 2, except that these are averages across the multiple years of earnings we observe (2008–2012). In Table 2 we use only the two pre-exit exam cohorts (2003–2004) to estimate these returns; this provides a useful benchmark for the results below. 2004 programs have higher returns to reputation than the other program groups; a ten percentile increase in college reputation is associated with a 14 percent increase in earnings for 2004 programs, but only a three percent increase for 2009 programs.[18]

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

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

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

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

where w_{ipct} is the log average daily earnings for student i in program p, graduation cohort c, and with potential labor market experience t, defined as calendar year minus graduation cohort. d_{pc} are dummies for program-cohort cells and $f_p(t)$ is a quadratic in experience

[18] The negative return to reputation for the 2006 program illustrates the empirical challenge of trying to estimate a return to reputation within each program. Not only can these returns be noisy when only a few schools offer a program, but the value of going to a higher-ranked school depends on the labor market that students from the program commonly enter (in this case, the program trains surgical instruments technicians). For related issues see Hastings et al. (2013) and Urzua et al. (2015).

interacted with program dummies. This "first-step" specification estimates returns to college reputation, r_{pc}, and to ability, a_{pc}, separately for each program-cohort cell.

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

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

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

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

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

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

Proposition 1 predicts $\beta^r < 0$ and $\beta^a > 0$. This comes from the assumption that employers use both labor market reputation, $\mathscr{R}_s$, and other signals of worker skill, y_i, in setting initial wages. We assume that the exit exam increases the precision of y_i, for example, through the appearance of scores on CVs. Our measure of reputation, R_s, is a better predictor of $\mathscr{R}_s$, while Icfes scores, τ_i, are a better predictor of y_i. Thus as the market relies less on $\mathscr{R}_s$ and more on y_i, the return to reputation falls ($\beta^r < 0$) and the return to ability rises ($\beta^a > 0$).[19]

Column (A) of Table 3 estimates benchmark specification (8). Like all other columns in Table 3 it reports only the β^r and β^a coefficients on the interactions of reputation and Icfes with our treatment variable δ_{pc}. The results suggest that relative to students in programs and cohorts without exams, students exposed to the exit exams see their daily earnings become more correlated with incoming collegiate ability and less correlated with college reputation.

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

The reputation effect is slightly lower than one third of the mean return to reputation in Table 2; the Icfes coefficient is slightly higher than one half of the mean return to Icfes.[20]

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

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

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

3.6.2. *Experience and cohort controls.* Our sample includes 2003–2009 cohorts with earnings measured in 2008–2012. This means we cannot disentangle a first-period effect of the exit exam from an effect that varies with experience because we do not observe initial earnings for pre-exit exam cohorts. As a result, our benchmark results are based on returns to reputation and ability that average across experience levels.

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

[20] Appendix B.5 presents the program-cohort level returns to reputation and Icfes from the first-step equation (6). Averaging and differencing these returns yields estimates similar to column (A) of Table 3.

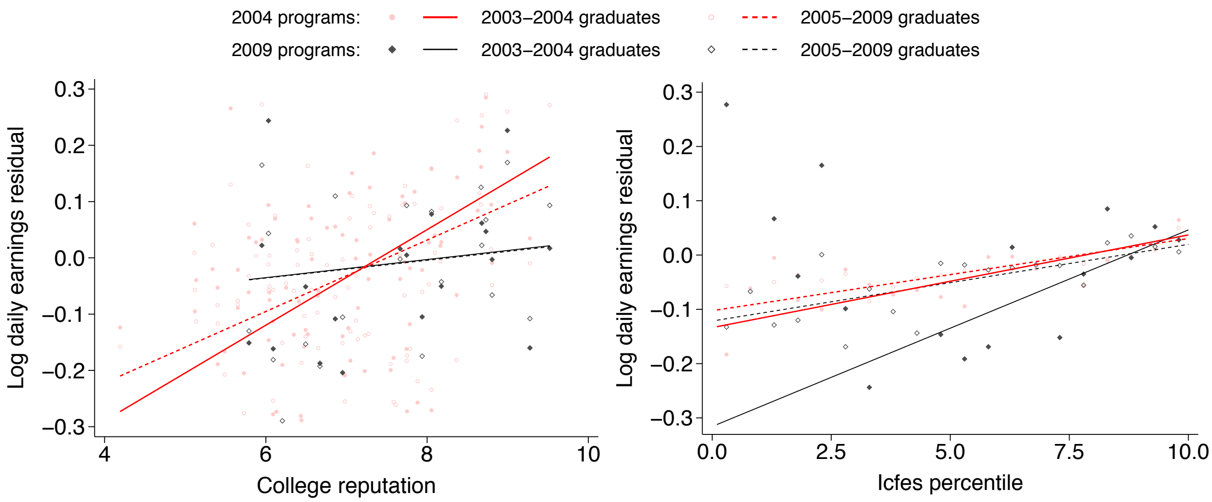
Table 3. Exit exam effects on returns to reputation and ability
Dependent variable: log average daily earnings

	(A)	(B)	(C)	(D)	(E)	(F)
		Experience & cohort controls		Restriction to similar programs		
	Benchmark specification	Within experience	Linear trends	S. sciences & engineering	Within $\hat{r}_p$ quartiles	Within $\hat{a}_p$ quartiles
Reputation × δ_{pc}	−0.041**	−0.033**	−0.034	−0.046*	−0.026***	−0.051***
	(0.017)	(0.015)	(0.028)	(0.026)	(0.010)	(0.017)
Icfes × δ_{pc}	0.017***	0.018**	0.012	0.038***	0.016***	0.017***
	(0.006)	(0.007)	(0.009)	(0.010)	(0.005)	(0.005)
N	581,802	267,924	267,924	273,590	581,802	581,802
R^2	0.258	0.224	0.224	0.266	0.258	0.258
# programs	39	39	39	22	39	39
Experience levels	0–9	4–7	4–7	0–9	0–9	0–9

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

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

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



Panel A. Return to reputation

Panel B. Return to ability

FIGURE 3. Exit exam effects—2004 and 2009 programs

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

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

To address this we add further controls for experience to the benchmark specification. To illustrate, suppose we estimated (8) using only earnings at five years of potential experience, thus ensuring that we are comparing exposed and unexposed cohorts at the same seniority. This regression could only include 2003–2007 cohorts because we do not observe earnings five years out for 2008–2009 graduates. We could repeat this estimation for any level of potential experience at which we observe cohorts prior to the introduction of all exit exams, which is between four (using 2004–2008 graduates) and seven (using 2003–2005 graduates) years of experience.[21] This procedure would yield four college reputation treatment effects and four Icfes treatment effects, one for each year of potential experience. We combine these into a single estimate by removing the experience quadratics from (8), restricting observations to those between four and seven years of experience, and fully interacting all fixed effects with

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

experience dummies:

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

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

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

A related test is to allow the returns to reputation and ability to follow program-specific linear trends in both experience t and cohort c. For this we add linear trend interactions with reputation ($\mu_p t R_s$ and $\mu_p c R_s$) and with Icfes ($\nu_p t \tau_i$ and $\nu_p c \tau_i$) to the benchmark specification.[22] Including experience trends alone yields similar estimates to those from specification (9) since we limit the sample to earnings between four and seven years of experience. Adding cohort trends is the typical differences in differences test of adding linear terms in the "time" dimension. Cohort trends absorb linear program-specific paths in the returns to reputation and ability that predate the exit exam and should have a measurable impact on our point estimates if these paths are important.[23]

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

3.6.3. *Restriction to similar programs.* Our key identifying assumption is that in the absence of the exit exams, there would have been parallel trends in the returns to reputation and ability among programs exposed and not exposed to the exams. One fact that might cast doubt on this is that programs that got exams early have higher returns to reputation (Table

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

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

[23] Our ability to control for pre-existing cohort trends is limited, however, because we only observe two cohorts prior to the exit exam introduction (2003–2004).

2). To address this we focus on comparable programs. We do so in three ways: i) restricting attention to social sciences and engineering, areas that have multiple programs in different exam year groups (see Table 1);[24] ii) stratifying programs by quartiles of the pre-exit exam returns to reputation, and iii) stratifying programs by quartiles of the pre-exam returns to Icfes. In each case we define program groups G and supplement equation (8) with dummies for group-cohort cells interacted with reputation and Icfes (e.g., $\mu_{Gc}R_s$ and $\nu_{Gc}\tau_i$).[25] Thus, β^r and β^a are only identified by variation in exposure to the exit exam within groups of programs that have common characteristics.

Column (D) in Table 3 uses only programs in social sciences and engineering. The reputation effect is similar in magnitude to those in previous columns, while the Icfes effect is more than double. Both are statistically significant at the ten percent level despite the fact that the program restriction substantially reduces precision.[26]

In column (E) we define program groups by pre-exit exam returns to reputation. We first estimate a return to reputation for each of the 39 programs in our sample using 2003–2004 graduates (i.e., $\hat{r}_{p,2003-2004}$).[27] We then define program groups G by quartiles of these returns, with 9–10 programs per group. This directly addresses the concern that 2004 programs have higher returns to reputation—in this case we compare delayed exam programs with low reputation returns only to the subset of 2004 programs with similarly low returns. The reputation effect in column (E) is smaller than in earlier specifications, consistent with some inflation in our estimates due to pre-treatment differences; but it is still significant because the standard error decreases. This suggests that the effects in this specification are identified off more similar programs because there is less noise in estimating treatment effects.

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

3.6.4. *Placebo and balance tests.* A further placebo test replicates our main analysis using college *drop-outs* rather than graduates. Drop-outs are a compelling placebo group because

[24] The health program area also includes a single program with a delayed exit exam field (surgical tools). Estimates analogous to column (D) that include health programs yield similar coefficients, but they are not significant because identification in the health program area comes from this single program.

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

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

[26] We note, however, that column (D) of Table 3 does not adjust standard errors to account for the reduced number of program clusters, which is below the rule of thumb suggested by Angrist and Pischke (2009).

[27] For this we estimate equation (6) using only 2003–2004 graduates and replace the r_{pc} and a_{pc} coefficients with r_p and a_p. Appendix B.5 presents these program-specific returns to reputation (and returns to ability).

Table 4. Placebo test using college drop-outs

	(A)	(B)	(C)	(D)
	Dependent variable: Took the exit exam		Dependent variable: Log average daily earnings	
	Graduates	Drop-outs	Graduates	Drop-outs
Exposed to exit exam (δ_{pc})	0.500***	0.025		
	(0.054)	(0.020)		
Reputation $\times$ δ_{pc}			-0.041**	0.011
			(0.017)	(0.032)
Icfes $\times$ δ_{pc}			0.017***	-0.002
			(0.006)	(0.011)
N	146,052	77,586	581,802	259,258
R^2	0.335	0.026	0.258	0.118
# programs	39	39	39	39

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

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

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

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

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

they enroll in the same colleges and programs as graduates but exhibited little change in exam taking. Columns (A) and (B) in Table 4 document this by regressing an indicator for taking the exit exam on program dummies, cohort dummies, and our treatment variable, δ_{pc}. For graduates, exposure to the exit exam is associated with a 50 percentage point increase in the likelihood of taking the exam; for drop-outs it is unrelated.

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

This placebo test is consistent with balance regressions, reported in Appendix B.6, that ask whether the exit exam rollout was correlated with changes in graduates' observable characteristics. If the field-specific introduction of the exit exams were correlated with trends in school or program choice, this should appear as changes in average reputation or Icfes scores across programs. There is little evidence of this channel. Changes in reputation and Icfes scores in programs with access to the exit exams are small and statistically insignificant.[28]

Appendix B.6 also explores the effect of the exit exams on the probability of formal employment—a potential sample selection concern since we do not observe earnings for non-employed or informal workers. The estimated effect is not statistically significant and small relative to the mean formal employment rate.

3.6.5. *Other reputation measures and competing hypotheses.* Our measure of reputation, R_s, captures the expected "admission exam" ability of graduates from a given college. The exit exams may also have provided information to employers on other dimensions of graduates' skill. Table 5 explores some of these. Columns (A)-(C) present results that use different measures of college reputation but are otherwise identical to our benchmark specification (Table 3, column (A)). Column (A) defines reputation as mean Icfes at the college-program level rather than the college level, which allows schools to have strengths that vary by major. This is relevant because Colombian students apply to college/major pairs. Column (B) defines a college's reputation as one minus its admission rate (this measure is thus positively correlated with R_s). Column (C) defines reputation as the average log earnings of a college's graduates.[29] This yields our best measure of labor market reputation, $\mathscr{R}_s$, which includes both pre-college ability, α_i, and attributes related to college membership, v_s.

The exit exams led to an increasing return to Icfes and a lower return to reputation by all three measures, though the effect on the average earnings measure is statistically insignificant. The similarity of these results reflects the fact that R_s is mechanically correlated with other desirable school attributes when colleges use admission scores to select students. Reputation measures like average earnings do not provide a clean test of signaling, however, because they may be correlated with ability even conditional on individual Icfes scores.

Table 5 also illustrates how we can distinguish a signaling channel from other competing hypotheses. For example, the exit exams may have led to a declining influence of R_s because school-mean exit exam scores were publicized, potentially altering the market's perception of college reputation. The exit exams may also have prompted colleges to change curricula

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

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

Table 5. Exit exam effects under other reputation measures and hypotheses
Dependent variable: log average daily earnings

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

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

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

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

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

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

or add test-preparation sessions. If such reforms affected students' skill acquisition in a way that is positively correlated with pre-college ability, the exit exam introduction would increase the unconditional return to reputation or to individual Icfes scores.

Table 5 address these hypotheses by replicating our benchmark specification with only reputation terms (column (D)) or only Icfes terms (column (E)) included. The signs on both of these unconditional effects match our main results, but neither effect is statistically significant. This does not rule out any of the above hypotheses, as there may be offsetting effects. It does illustrate, however, that the strongest empirical result is the shift in weight from a group-level measure of ability—reputation—to an individual measure—Icfes scores. This is the effect captured by our benchmark specification, and it is harder to explain through channels other than signaling. Further, any institutional responses to the exams would affect attributes related to college membership—the v_s term in our model—and not pre-college ability, α_i, which is the focus of our signaling analysis.

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

3.7. **Complementary effects of the exit exam.** There is suggestive evidence that the exit exam affected other outcomes. For example, Column (A) in Table 6 shows its impact on time to graduation. This estimate is from a standard differences in differences regression that includes program dummies, cohort dummies, and our treatment variable, δ_{pc}. The result suggests that individuals in programs with exam fields took about one quarter of a year longer to graduate. This is consistent with increased student effort, or with colleges taking steps to prepare students for the test. For instance, there is anecdotal evidence of colleges seeking to influence their students' performance, with activities ranging from "boot camp" preparation to more overt "gaming" via exclusion of certain students.[30]

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

Finally, we ask whether the exit exams altered individuals' school or program choices. This would be consistent with the government's stated intent. Column (C) explores how the ability of incoming students changed with the exit exam introduction. For this regression we define two measures of reputation using a population of graduates who took the exit exam in 2009–2011, when it was required of all graduates. We define Icfes reputation as mean Icfes percentile at the *school-program* level. Similarly, exit exam reputation is the school-program mean exit exam percentile. We convert Icfes and exit exam scores to percentiles within this population so that both reputation measures are on the same scale.

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

Column (C) presents a specification analogous to the benchmark (8) with two key differences. First, the sample includes 2003–2009 *enrollees* rather than graduates, and we define

[30] These results suggest that graduation cohort may be endogenous. We address this concern by estimating (8) with cohorts defined by *predicted* rather than actual graduation date, where predicted graduation is based on the year of enrollment. Appendix B.4 shows that the results from this regression are similar to our benchmark specification; this suggests that selective graduation timing is not driving our main results.

Table 6. Complementary effects of the exit exam

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

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

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

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

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

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

students as treated by the exit exam ($\delta_{p\tilde{c}} = 1$) if they began a program p in an enrollment cohort $\tilde{c}$ after the introduction of the assigned field. Second, the dependent variable is the Icfes percentile of entering students, and we replace the independent variables R_s and τ_i with the school-program measures of Icfes and exit exam reputation. The reported coefficients in column (C) reflect how the correlations of Icfes and exit exam reputation with incoming students' Icfes scores changed with the exit exam rollout.[31]

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

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

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

4. College reputation and earnings growth

The previous section showed that college reputation plays a signaling role. This section asks whether college reputation serves *only* to signal ability as measured by admission scores. To do so, it tests Proposition 2 (Section 2), which predicts how college reputation correlates with initial earnings and with earnings growth.

4.1. **Sample.** We follow Farber and Gibbons (1996) and Altonji and Pierret (2001) in studying individuals making their initial transition to the labor force. We restrict our sample to individuals who: i) graduated in 2008 or 2009 (this allows us to observe earnings in the year of graduation and the next three years), and ii) entered the labor market immediately upon graduation and remained during four consecutive years (i.e., they did not attend graduate school or leave the formal labor force).[32] The results are thus not attributable to movements into and out of the labor market.

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

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

The dependent variable, w_{it}, is log daily earnings for student i measured at potential experience t, which as before is employment year minus graduation year; $d_{c_i t}$ are graduation cohort c_i by experience t cell dummies; college reputation, R_{s_i}, and Icfes score, τ_i, are as before; r_0 is the return to reputation in the year of graduation, and r is the average change in the return to reputation from an additional year of potential experience; a_0 is period-zero return to ability, and a is the average yearly change in this return.[33] We report only coefficients on reputation, Icfes, and their interactions with experience, where the latter two are estimated using earnings only up to three years after graduation, the maximum we can observe for our sample of 2008–2009 graduates.

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

[32] Appendix B.7 provides further details on the sample.

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

Table 7 estimates (10) both excluding and including Icfes terms, which yields the unconditional return to reputation and the conditional returns to reputation and Icfes. This corresponds to regressions (4) and (2) from Section 2 and the various subparts of Proposition 2.[34] We discuss results from each of these regressions separately in the subsections below.

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

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

The contrast between the reputation and years of schooling results can also be depicted using earnings-experience profiles. Mincer (1974) noted that the wage profiles of workers with different schooling levels are approximately parallel throughout the earnings lifecycle. Panel A of Figure 4 replicates this finding using 2008–2012 household survey data from Colombia.[35] It plots the mean log hourly real wage among workers with two schooling levels—completed high school and completed college—i.e., the gap between the two profiles is the college premium. This gap remains roughly constant across forty years of potential experience, consistent with results in the U.S. (Lemieux, 2006).[36]

Panel B uses our administrative data to plot earning profiles by college reputation. To match the cross-sectional analysis in Panel A, Panel B includes 2008–2012 earnings from all 2003–2012 college graduates. We plot mean log daily real earnings separately for graduates

[34] Proposition 2 also contains predictions for regressions that include Icfes but not reputation terms. Appendix B.8 shows that the results match the predictions: the unconditional return to Icfes increases with experience. This is consistent with findings in Farber and Gibbons (1996) and Altonji and Pierret (2001).

[35] In Figure 4, we define potential labor market experience as $\min(\text{age} - \text{years of schooling} - 6, \text{age} - 17)$. This definition differs from the one we use elsewhere in the paper (earnings year minus graduation year) because the Colombian household survey does not include school completion dates. However, the age and schooling definition matches those in Mincer's original analysis and in Altonji and Pierret (2001).

[36] The constant relationship between years of schooling and earnings in Colombia also holds in standard Mincerian regressions reported in Appendix B.9.

Table 7. Returns to reputation and ability, and experience interactions
Dependent variable: log average daily earnings

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

Notes: The dependent variable is log average daily earnings. The sample includes students in column (D) of Appendix Table B7 and earnings in the four years after graduation. Columns (A) and (B) estimate equation (10) excluding and including Icfes terms, respectively. In addition to the reported variables, both regressions include dummies for cohort-experience cells.

Column (C) adds the following controls to column (B): 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.

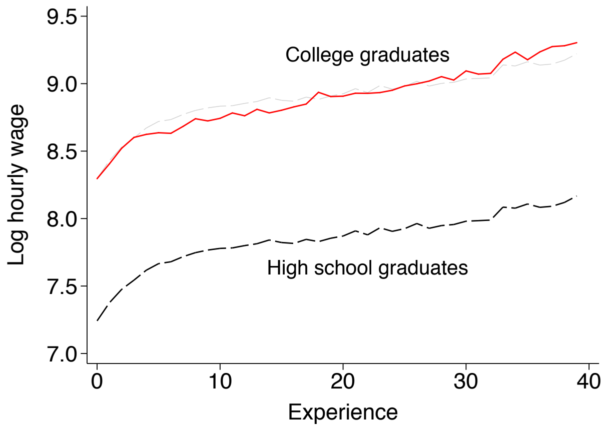
Parentheses contain standard errors clustered at the college level.

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

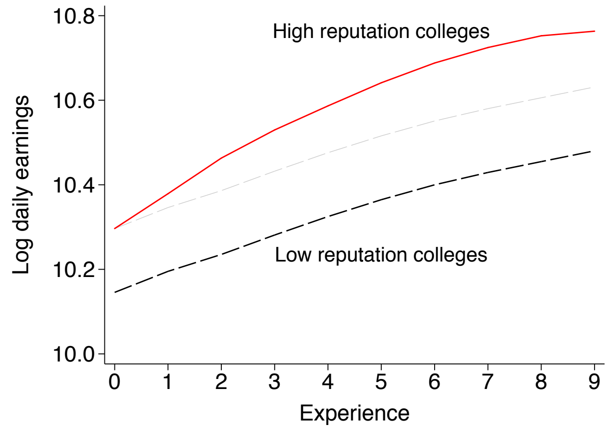
from high and low reputation colleges, defined by the median reputation. The earnings gap between the two profiles roughly doubles over the first ten years of experience, as indicated by the divergence of the high reputation profile from the light grey dashed line that is parallel to the low reputation profile.

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

4.2.2. *Conditional returns to reputation and ability.* Column (B) of Table 7 estimates equation (10) as written. In this joint specification, the coefficients reflect the conditional returns to reputation and to ability from equation (2). As Proposition 2 predicts, the period-zero



Panel A. Years of schooling



Panel B. College reputation

FIGURE 4. Earnings-experience profiles

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

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

reputation coefficient is lower than its unconditional return in column (A). Consistent with employer learning about ability, column (B) also shows a positive and significant coefficient on the interaction of Icfes and experience.[37]

The main coefficient of interest is on the interaction of reputation with experience. Proposition 2 states that the conditional return to reputation should fall over time. This is similar to the Altonji and Pierret (2001) prediction for observable traits like race or schooling, but our definition of reputation yields an even stronger test of signaling. Since reputation is a group-level mean of Icfes, Icfes scores are a sufficient statistic for "admission exam" ability,

[37] The positive coefficient on the Icfes-experience interaction is similar to the Farber and Gibbons (1996) and Altonji and Pierret (2001) findings using Armed Forces Qualification Test (AFQT) scores as an unobserved characteristic. However, it is in contrast with findings in Arcidiacono et al. (2010), who also study AFQT scores but make a distinction between graduates who enter the labor market after high school and those who do so after college. For college graduates, they show that AFQT is strongly related to wages in the year of graduation, and this relationship changes little over the next ten years. Their conclusion is that AFQT revelation is complete for college graduates, and they suggest that this revelation occurs through college identity. Appendix B.8 discusses one potential explanation for the difference in findings: sorting by ability in Colombia—although increasing—appears to be less extensive than in the U.S.

α_i; conditional returns to reputation *mechanically* do not reflect the transmission of information on α_i. The conditional return to reputation should, therefore, decline with experience even if employers do not perfectly observe our measure of reputation; learning about reputation is reflected in the Icfes coefficients. Unlike Altonji and Pierret (2001), our model predicts a negative coefficient on Reputation$\times t$ even if there are interactions between ability, α_i, and human capital growth, h_{it}. These effects are also captured by the Icfes$\times t$ term.

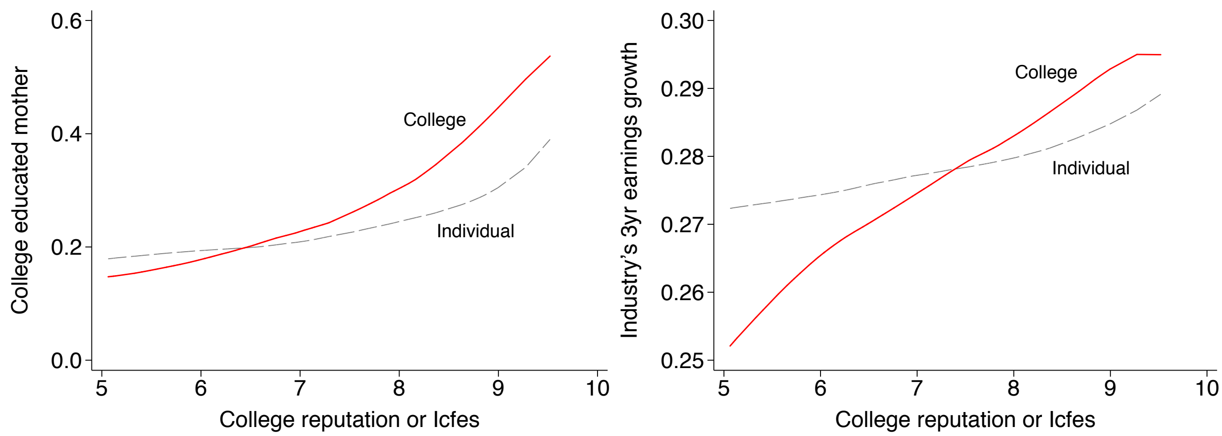
In sum, if college reputation serves purely as a signal of ability, Proposition 2 predicts a negative coefficient on the interaction of reputation and experience. Column (B) clearly rejects this. The reputation-experience interaction, although smaller in magnitude than in column (A), is still positive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

The increasing correlation of reputation and earnings is a descriptive result, but it is robust to a wide range of specifications and samples. For example, Column (C) of Table 7 adds controls for graduates' gender, age, socioeconomic status, college program, and regional market. All controls are interacted with a quadratic in potential experience to allow earnings trajectories to vary with each characteristic. The coefficient on the reputation-experience interaction decreases slightly, but it is still highly significant and roughly of the same magnitude. Appendix B.10 shows that this interaction term remains positive with further controls, different definitions of labor market experience, and in alternate samples.

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

Figure 5 provides suggestive evidence that both of these channels may be at work. First, Panel A considers one potential component of v_{s_i}: socioeconomic status as measured by whether a student's mother has a college degree.[38] The x-axis contains reputation when observations are colleges, and Icfes when observations are individuals (the scale is the same). The solid line shows that as one moves from the college with the lowest reputation to that with the highest, the mean percentage of students with college-educated mothers increases

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



Panel A. Socioeconomic status Panel B. Industry's earnings growth

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

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

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

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

from below 20 to above 50. The dashed line describes the individual-level relationship between students' Icfes scores and their mother's education, i.e., this is the relationship that would exist if sorting into colleges were by Icfes only. Socioeconomic sorting is less pronounced in this hypothetical scenario than in the actual one; i.e., there is more sorting across colleges on mother's schooling than is predicted by Icfes scores alone.[39] This is consistent with a positive reputation premium ($v_1 > 0$); sorting on mother's education is positively correlated with reputation. This could lead to a rising return to reputation if employers imperfectly observe both reputation and mother's education.

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

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

Table 8. Industries of first employment ranked by reputation residuals

Top 10 industries			Bottom 10 industries		
Industry	Mean reputation residual	3 year earnings growth	Industry	Mean reputation residual	3 year earnings growth
Securities trading	0.52	0.67	Public transportation	-0.45	0.20
Higher education	0.41	0.27	Financial cooperatives	-0.41	0.17
Oil and gas extraction	0.31	0.50	Social security services	-0.37	0.21
Specialized animal breeding	0.30	0.15	Preschool education	-0.34	0.22
Chemical manufacturing	0.28	0.52	Intermediate products wholesale	-0.33	0.19
Petroleum and natural gas extraction	0.28	0.47	Basic primary education	-0.32	0.19
Pharmaceutical products wholesale	0.27	0.42	Gambling services	-0.31	0.23
Rubber products manufacturing	0.27	0.20	Public services and administration	-0.30	0.08
Cleaning products manufacturing	0.26	0.45	Elementary education facilities	-0.29	0.17
Radio and television	0.26	0.54	Soda and mineral water production	-0.27	0.32
Top 10 average	0.33	0.36	Bottom 10 average	-0.34	0.19

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

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

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

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

These results suggest that the increasing return to reputation may reflect a career effect (Topel and Ward, 1992) in which better college reputation allows some individuals to be matched to jobs with steeper wage profiles, or to firms that facilitate more on-the-job training. Higher reputation schools might also provide better networks (e.g., Kaufmann et al., 2013; Zimmerman, 2013) that ultimately make individuals more productive.[40]

Our setting and data do not reveal whether the correlation between college reputation and earnings growth is due to unobserved dimensions of sorting or due to a causal effect of college identity. But the widening of earnings profiles across Colombian colleges is starkly different from the parallel nature of earnings profiles across schooling levels. This may lead students to suspect that their choice of college quality matters for their earnings trajectories in a way that their choice of educational attainment might not.

5. Conclusion

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

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

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

We showed that, consistent with work on other markets, employers use college reputation to make inferences about individual graduates. Specifically, while the cross-sectional data are consistent with this, we exploited a natural experiment in Colombia to show that providing

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

more information about student skill reduces the importance of reputation. Thus college identity performs a signaling function, and students may be right to worry about *which* college in addition to *whether* college. In other words, we find support for MacLeod and Urquiola's (2015) assumption that labor markets do not immediately observe all individual characteristics (such as Icfes or AFQT scores), and college membership may transmit some of them.

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

The purpose of the big sort is to match individuals to jobs. A literature documents significant differences in compensation across firms and occupations that cannot be explained by worker ability (Krueger and Summers, 1988; Gibbons and Katz, 1991; Abowd et al., 1999). Our results are consistent with a set of social norms in which the labor market allocates the "better" jobs to individuals from more reputable schools.

References

Abdulkadiroglu, A. and T. Sonmez (2013). Matching markets: Theory and practice. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics*, Volume Vol I: Economic Theory, Journal article 1, pp. 3–47. Cambridge University Press.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333. 00129682 Accession Number: 0488247; Keywords: Pay; Wage; Geographic Descriptors: France; Europe; Geographic Region: Europe; Publication Type: Journal Article; Update Code: 199905.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2014). Trends in earnings differentials across college majors and the changing task composition of jobs. *American Economic Review 104*(5), 387–393.

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

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

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

Arcidiacono, P. (2004, JUL-AUG). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1-2), 343–375.

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

Avery, C. N., M. E. Glickman, C. M. Hoxby, and A. Metrick (2013). A revealed preference ranking of us colleges and universities. *Quarterly Journal of Economics 128*(1), 425–467. Avery, Christopher N. Glickman, Mark E. Hoxby, Caroline M. Metrick, Andrew.

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

Black, D. A. and J. A. Smith (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics 24*(3), 701–728.

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

Cabral, L. and A. Hortaçsu (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics 58*(1), 54–78.

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

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

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

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

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

Dranove, D. and G. Z. Jin (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature 48*(4), 935–963.

Epple, D., R. Romano, and H. Sieg (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica 74*(4), 885–928.

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

Gibbons, R. and L. F. Katz (1991). Layoffs and lemons. *Journal of Labor Economics 9*(4), 351–80.

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

Grogger, J. and E. Eide (1995). Changes in college skills and the rise in the college wage premium. *The Journal of Human Resources 30*(2), pp.280–310.

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

He, Y. (2014, November). Gaming the boston school choice mechanism in beijing. Mimeo, Toulouse School of Economics.

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

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

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

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

Hoxby, C. and C. Avery (2013). The missing 'one-offs': The hidden supply of high-achieving, low income students. *Brookings Papers on Economic Activity Spring*, 1–65.

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

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

Hubbard, T. N. (2002, Oct). How do consumers motivate experts? Reputational incentives in an auto repair market. *Journal of Law & Economics 45*(2), 437–468.

Jin, G. Z. and P. Leslie (2003). The effect of information on product quality. *Quarterly Journal of Economics 118*(2), 409–451.

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

Kahn, L. B. and F. Lange (2014). Employer learning, productivity, and the earnings distribution: Evidence from performance measures. *Review of Economic Studies 81*(4), 1575–1613.

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

Krueger, A. B. and L. H. Summers (1988, March). Efficiency wages and the inter-industry wage structure. *Econometrica 56*(2), 259–294.

Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics 25*(1-35), 497–532.

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

MacLeod, W. B. and M. Urquiola (2015). Reputation and school competition. *American Economic Review forthcoming*.

Melnik, M. I. and J. Alm (2002, September). Does a seller's ecommerce reputation matter? Evidence from ebay auctions. *Journal of Industrial Economics 50*(3), 337–49.

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

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

Roth, A. E. and M. Sotomayor (1989). The college admissions problem revisited. *Econometrica 57*(3), 559–570.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics 98*(4), 659–79.

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

Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality. *Review of Economic Studies 63*(1).

Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics 107*(2), 439–479.

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

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