Wp

EDUCATION, SIGNALING AND MISMATCH

Arozamena, Leandro
Ruffo, Hernán

CAF DEVELOPMENT BANK OF LATIN AMERICA

EDUCATION, SIGNALING AND MISMATCH

Arozamena, Leandro

Ruffo, Hernán

ABSTRACT

We assess the importance education as a signal of workers skills and the e¤ects of poor signaling quality on labor market outcomes. We do so by merging a frictional labor market model with a signaling setup where there is a privately observed idiosyncratic component in the cost of education. Given that highly skilled workers cannot correctly signal their abilities, their wages will be lower and they will not be matched to the "right" vacancies, or may be unemployed. Skilled workers will then have lower incentives to move to high productivity markets. Furthermore, fewer vacancies will be created in labor markets where skills matter, and incentives for workers to invest in education will be lower. Overall, an economy where education is a noisier signal generates lower educational attainment, higher unemployment and lower productivity. In addition, we provide evidence suggesting that education plays a poor signaling role in Latin American countries. We then calibrate our model using Peruvian data, and through a quantitative exercise we show that this mechanism could be relevant to explain the relatively bad performance of labor markets in Latin American countries.

EDUCACIÓN, SEÑALIZACIÓN Y DESAJUSTE

Arozamena, Leandro
Ruffo, Hernán

RESUMEN

Evaluamos la importancia de la educación como mecanismo de señalización de las habilidades de los trabajadores, y los efectos de una mala señal sobre resultados laborales. Lo hacemos mediante la fusión de un modelo de mercado laboral con fricciones con un modelo de señalización que incorpora un componente idiosincrásico en el costo de la educación, observado de forma privada. Dado que los trabajadores altamente calificados no pueden señalar correctamente sus habilidades, sus salarios serán más bajos y no cubrirán las vacantes "correctas", o permanecerán desempleados. Los trabajadores calificados tendrán entonces menos incentivos de moverse a mercados de alta productividad. Por otra parte, un menor número de vacantes será creado en aquellos mercados de trabajo donde las habilidades importan y los incentivos para los trabajadores de invertir en la educación serán menores. En general, una economía en que la educación es una señal ruidosa genera menor nivel educativo, mayor desempleo y menor productividad. Además, proporcionamos evidencia que sugiere que la educación juega un papel mediocre en la señalización en América Latina. Calibramos nuestro modelo usando datos de Perú, y a través de un ejercicio cuantitativo se muestra que este mecanismo podría ser relevante para explicar el relativamente mal desempeño de los mercados laborales en América Latina.

Education, Signaling and Mismatch

Leandro Arozamena and Hernán Ruffo[1]
Universidad Torcuato Di Tella

March 2016

Abstract   We assess the importance education as a signal of workers' skills and the effects of poor signaling quality on labor market outcomes. We do so by merging a frictional labor market model with a signaling setup where there is a privately observed idiosyncratic component in the cost of education. Given that highly skilled workers cannot correctly signal their abilities, their wages will be lower and they will not be matched to the "right" vacancies, or may be unemployed. Skilled workers will then have lower incentives to move to high productivity markets. Furthermore, fewer vacancies will be created in labor markets where skills matter, and incentives for workers to invest in education will be lower. Overall, an economy where education is a noisier signal generates lower educational attainment, higher unemployment and lower productivity. In addition, we provide evidence suggesting that education plays a poor signaling role in Latin American countries. We then calibrate our model using Peruvian data, and through a quantitative exercise we show that this mechanism could be relevant to explain the relatively bad performance of labor markets in Latin American countries.

1   Introduction

The role of education and its impact on labor market performance and job productivity has long been recognized as key. Both human capital accumulation and signaling of workers' abilities have been identified as channels through which education plays this role. This paper concentrates on one specific aspect of the signaling channel: signal quality. Specifically, we study the implications of education as a (possibly poor) signaling technology on allocation and productivity in labor market equilibrium. If education provides less precise information about workers' abilities, an allocation problem follows. At least initially, matching worker skills with the right vacancies is problematic. Unless the labor market is frictionless, inefficient worker allocation could be long lasting, resulting in lower

---

overall productivity. Here, we intend to examine this problem both qualitatively and quantitatively. In particular, we aim to study to what extent this specific channel could contribute to explain the relatively poor performance of labor markets in Latin American countries.

First, we build a model combining two contexts that have been previously examined in the literature: a general model of mismatch between workers and vacancies in different labor markets within an economy, based on Shimer (2007), and a signaling model where education may or may not convey precise information about a worker's abilities. We extend the basic mismatch model to introduce the decision of workers to reallocate. We modify the standard, pure signaling setting by adding an idiosyncratic shock to the cost workers face when educating. This heterogeneous cost is intended to reflect that a given worker's ability will not necessarily be easy to signal through education.

Indeed, this heterogeneity implies that some highly skilled workers optimally decide not to invest in education. Given that these workers cannot correctly signal their abilities, their wages will be lower and they will not be matched to the "right" vacancies. Misallocation follows through several mechanisms. Employers cannot correctly identify worker skills, so some skilled workers will be unemployed. Skilled workers will then have lower incentives to move to high productivity markets. Furthermore, fewer vacancies tend to be created in labor markets where skills matter, and incentives for workers to invest in education tend to fall. On the hole, an economy where education is a noisier signal generates lower educational attainment, higher unemployment and lower productivity.

In addition to examining those effects qualitatively through our model, we attempt to quantify them. Our quantitative approach intends to explore the relevance of the signaling mechanism for explaining a set of facts present in Latin American labor markets. First, Latin American countries have relatively bad quality of education, as measured by standardized tests. As Hanushek and Woessman (2012) have documented, those who have stayed at school until age 15 in the region perform badly in terms of cognitive skills relative to their peers in OECD countries. Second, the average number of years of education of workers is lower in Latin America than in developed countries. Third, worker misallocation is prevalent in Latin American countries (evidenced in a number of indicators, such as informality, size distribution of firms, or unemployment), which yields lower productivity (Erosa and Allub, 2012, D'Erasmo and Moscoso Boedo, 2012, Buera et al., 2011). Latin American countries are then a natural choice to examine to what extent poor quality of education as a signal of worker skills has an impact on labor market performance. Indeed, we will show that lower signal quality of education in Latin American countries can account for those three facts.

To do so, we first provide and discuss the evidence that allows us to conclude that education plays a poorer signalling role in Latin America than it does in OECD countries, resorting to a number of available skills measures. In Latin America, there is a larger mismatch between the skills required in different occupations and the skills those employed in those occupations actually have.

Furthermore, the variance of skills indicators is larger for each educational level, suggesting that education does not perform well as a signal. That performance is even worse in the case of noncognitive than for cognitive skills. Regression analysis yields that wages are correlated not only with skills indicators, but also with education levels once skills are controlled for. In our model, this is only consistent with education being a bad signal of skills. In the paper we also discuss up to what extent measurement error in skills variables could be an issue for the link between the data and our model.

Then, we calibrate the model for Peru using available data on education, cognitive and noncognitive skills by occupation. Then, by changing parameters related to signaling, we assess the quantitative importance of signaling and the mismatch process, analyzing the productivity effects that arise due to a worse signaling technology. Indeed, poorer signal quality generates lower productivity. For example, one of our quantitative exercises yields a productivity drop of 24% resulting from a poor signal -as compared to the case where workers' abilities are perfectly revealed through education- due to human capital misallocation

After reviewing the related literature, the paper follows by describing our framework in detail in Section 2. In Section 3, we provide a simplified model to understand the economic mechanism through which signal quality affects the economy. In Section 4, we describe the data that provides us with relevant estimations to calibrate our model. We describe the calibration procedure in Section 5. Then, we use our calibrated economy to derive results in Section 6 and discuss our findings in Section 7. Finally, Section 8 offers some conclusions.

Related literature   This paper is related to several strands of previous work. Ever since Spence (1973), a large literature has examined both theoretically the potential signaling role of education (Riley, 2001, provides a survey). We modify the standard signaling setting by adding an idiosyncratic shock to the cost of education, so that education is an imperfect signal and the effects of signal quality can be examined.[2]

The standard signaling approach has been changed in different ways so as to incorporate the difficulties that may limit education's role as a signal. Feltovich et al. (2002), for example, take the case where an alternative measure of a worker's abilities (e.g. a recommendation) is available in addition to education. Countersignaling may follow: education may be selected nonmonotonically according to skills. If the alternative measure is very informative, some worker types may not be inclined to select higher education levels to transmit information on their abilities to firms. Araujo et al. (2007) develop a model of "mixed signals", in which both cognitive and noncognitive skills are private information. If there is a large divergence in how those two kinds of skills enter into the cost of education for workers and into firms' technologies, countersignaling may arise. Multidimensional private information may, then, make it difficult

---

[2] An alternative approach, adopted in the signaling models in Matthews and Mirman (1983), Carlsson and Dasgupta (1997) and de Haan et al. (2011), would be to add noise to the signal's value. In our application to education, we consider a cost shock more natural.

for education to convey precise information about skills. In our approach, we change the standard setup so that signal quality is associated with a shock to the cost a worker faces when educating. We can then use the variance of that shock as a measure of signal quality.[3]

Signaling theory has been empirically tested by a large literature. While sharp results are difficult to find, many papers present important evidence in favor of the signaling hypothesis (Tyler et al. 2000; Weiss, 1995). Under the idea that the signal conveyed by education is useful early in workers' labor history, a set of papers analyze the life-cycle earnings profile (Cohn et al., 1987; Layard and Psachaporoulus, 1974; Wolpin, 1977). Another group of papers compare returns to education in different labor markets with diverse signaling requirements. For example, several papers compare returns to education for wage earners and for the self-employed, since signaling should be irrelevant for the latter (Cohn et al., 1987; Katz and Ziderman, 1980; Riley, 1979). A third branch of the literature examines returns to credentials rather than to years of education (Groot and Oosterbeek, 1994; Jaeger and Page, 1996), and find that credentials have returns above the number of years of education. In a particularly interesting contribution, Tyler et al. (2000) examine the General Educational Development (GED) exam, usually taken by high-school dropouts in the US. The authors show that the GED credential has strong effects, between 10 to 19% higher wages, once controlling for GED scores.

Instead of concentrating on the relationship between education and labor income, other papers focus on workers' education decisions. As an example, Bedard (2001) compares high school dropout rates in locations with varying degrees of university access. Lang and Kropp (1986) study changes in enrollment rates under different compulsory school attendance legal requirements.

Still, all this empirical literature has not used information on actual skills to examine the relevance of signaling. Here, on the contrary, we intend to make use of available skills measures, which we will later describe. This information will be key in our quantitative analysis.

In addition to signaling, another strand of related literature examines different sources of misallocation, as surveyed in Hopenhayn (2014). As is emphasized in that paper, in order to quantitatively account for productivity gaps between countries, market distortions should generate ranking reversals in the size distribution of firms or in the use of factors, so that highly productive firms or factors end up unexploited. Our signaling model yields this type of misallocation: highly skilled workers can be unemployed while unskilled workers are not.

The issue of occupational mismatch has also been examined empirically. It is usually measured by comparing the skills required for each job with the skills that workers actually have. Following different approaches, Sicherman (1991), Chevallier (2003), Pellizzari and Fichen (2013), Allen and van der Velden (2001) and Desjardins and Rubenson (2011) provide attempts to measure mismatch.

---

[3]MacLeod et al. (2015) argue that, besides the number of years of education, the particular college that a worker attended provides some signal that employers might use to statistically discriminate among workers. They find empirical support for their observation.

We will describe these contributions in more detail when presenting our own approach, in Section 4.

In addition to cognitive skills, we also consider the role of noncognitive skills in the empirical part of the paper. A broad empirical literature has shown that both cognitive and noncognitive skills matter significantly to firms, labor demand and wages. For example, Heckman and Kautz (2012) illustrate how low noncognitive abilities reduce survival rates in employment, wages and annual labor income. How both types of abilities are acquired by workers and identified by actual and potential employers is a key issue for productivity and efficiency in labor markets. Noncognitive abilities may be harder to accumulate and to signal through formal education. Kautz et al. (2014), for instance, argue that the GED tends to send a mixed signal to potential employers. Those that pass the GED exam have higher cognitive abilities as compared with high-school dropouts that do not pass the test, but lower noncognitive abilities than those that have graduated from high school. Bassi et al. (2012) show that in Latin American countries a human resources policy in firms can be described as "recruit for attitude and train for skill." They connect this with the importance of noncognitive skills.

In any case, to the best of our knowledge, the impact of education as a signal on labor market outcomes has not been studied by combining it with a frictional labor market model. This, then, is this paper's main contribution.

2 A mismatch model with signaling

The model may be described as one combining two potentially distinct parts: (i) a dynamic model of mismatch in labor markets, and (ii) a simplified version of a signaling model for education. We introduce both parts in a series of steps.

2.1 Basic setup

There is a mass L of labor markets in the economy, which we normalize to one. Each labor market may be interpreted as a local market for an occupation. There are K types of markets or occupations, and L_k is the mass of type$-k$ markets, $k = 1, ..., K$. Markets may differ according to how important a worker's ability is for productivity: some occupations may be more skill-intensive than others. Therefore, workers with the same characteristics -as observed by firms- will receive different wages depending on the market they are employed in. Let N_k be the number jobs per market available in markets of type k, which will be determined endogenously, and let N be the total number of jobs in the economy.

There are M risk neutral workers. A worker's ability is given by ψ, where $\psi \in \Psi$. The set Ψ is finite, for simplicity. Each worker's utility will equal the expected present value of his income flow, discounted at rate r, minus the cost of education, as we will explain shortly.

Initially, each worker is assigned to a market at random. There, if a vacancy is available to him, he can produce at a job or stay unemployed. If employed at

market k, the worker will produce units of the consumption good according to

$$y_k(\psi) = A_k \psi^{\alpha_k}.$$

Then, a worker's productivity in a given market depends on his own abilities and, as we mentioned before, on the market type. If the worker remains unemployed, he produces z units of the consumption good at home.

Worker's abilities are not observable to employers. Before being initially assigned to a market, each worker chooses an education level e, observed by firms, which may play the role of a signal for his ability. Note that, for the sake of simplicity, given our interest in the effects of its role as a signal, we are assuming that education has no impact on actual productivity. Potential employers can compute the expected productivity of any worker that has chosen any given education level. Let $p_k(e)$ be the expected productivity, in a type$-k$ market, of a worker that has chosen an education level e. That is,

$$p_k(e) = E_\psi \left[A_k \psi^{\alpha_k} | e \right].$$

In what follows, we first describe how the assignment of workers to markets and the evolution of vacancies in each market work, given how workers have chosen their educational levels. Once we have described what workers could expect about employment and wages, we will come back to how their educational choices are initially made.

2.2 Flows

As mentioned above, each worker is initially assigned at random to a market. We assume that workers can be reallocated both randomly and voluntarily. Let q be the constant rate at which workers are randomly reallocated, which we will refer to as the "quit" shock. This rate is common to all markets and worker types and is independent of the worker's employment status. Any worker may also decide to abandon his current market for a new one, after paying a fixed cost Γ.

Both if he is affected by the quit shock and if he voluntarily leaves his current market, the worker is assigned to a new market at random. He may end up at any market with equal probability, independently of his characteristics, previous market allocation and past employment. Given that this probability is independent across workers, the allocation of workers across labor markets is a multinomial random variable.

Additionally, there is a constant rate l at which any given job is destroyed. If a job is destroyed, it is lost. A new job created in any market type is randomly allocated to a market within that type. There is no aggregate shock in the economy.

Let M_e be the number of workers that have selected education level e, and let $\mathbf{i} = (i_e)_{e=e_1}^{e_S}$ be the the vector that lists the number of workers in a given market for each education level e.[4] In a market where workers are described by $\mathbf{i}$, the

---

[4] We assume here that the number of education levels selected by workers, S, is finite, which will be the case later when we describe the signaling equilibria we will examine.

number of workers exogenously reallocated is distributed binomial, $B(\mathbf{i}, q)$, with mean $q \sum_e i_e$. Let j denote the number of jobs available in a market. Then, the number of jobs that are exogenously destroyed is distributed binomial, $B(j, l)$, with mean lj. The inflows to each market due to exogenous reallocation shocks have multinomial distributions, and could be approximated by their Poisson limits, $Poiss(qM_e)$ ($Poiss(lN_k)$) for workers with education e (respectively, for jobs in markets of type k). We can compute the processes of $\mathbf{i}$ and j generated by these exogenous shocks, as well as the stationary distributions for these variables in each market.

However, workers' endogenous reallocation decisions generate additional flows. The situation faced by a worker of eduation e in a market k will be fully described by the combination of workers in the market, $\mathbf{i}$, and jobs, j. Let $\varphi_e^k(\mathbf{i}, j)$ be a worker's policy function in a market of type k when his education level is e. It takes a value of one (zero) if the worker decides to switch markets (respectively, stay). These policy functions will, of course, follow from solving the worker's optimization problem. Let $\widehat{M}_e$ be the mass of workers with education e that decide to switch. The inflow of workers to each market should also include the corresponding $Poiss(\widehat{M}_e)$ distributions.

Considering both exogenous and endogenous entries and exits in each market, we can compute the joint distribution of workers $\mathbf{i}$ and jobs j for each market type k. Let $G_k(\mathbf{i}, j)$ be the cdf of this distribution.

For future use, let $i = \sum_e i_e$ be the number of workers in a given market, and let $h \equiv j - i$. If $h > 0$ ($h < 0$), the market is tight, in the sense that there are unfilled vacancies (respectively, there is unemployment).

2.3   Wages and value functions

There is perfect competition within each market. Furthermore, we will assume that workers are hired in spot markets -i.e. it is not possible to generate long-term relationships or contracts between any firm and any worker. The only relevant information about the worker that is available to a potential employer is his education level. Let $w_k(e)$ be the wage received by workers with education e in a given market of type k. Clearly, this wage depends on the number of workers of each type in the market, the number of jobs available and the market type.

To establish wage levels in a type-k market given the numbers of jobs and workers, let us start by examining the case where only workers with the same education level e have been allocated to that market. If $h < 0$ -that is, if there is unemployment- competition implies that the wage level will be given by worker's opportunity cost, i.e. $w_k(e) = z$. If, on the contrary, $h > 0$, there will be unfulfilled vacancies and firms will compete so that $w_k(e) = p_k(e)$: wages will equal expected productivity. We assume that firms will pay $w_k(e) = z$ if $h = 0$ as well.

Take now the case of a market where workers with different education levels coexist. Suppose those education levels are $e_1, ..., e_S$, with $e_{s+1} > e_s$, $s = 1, ..., S$. If $h > 0$, all workers will receive a wage that equals their expected

productivity, according to each education level. That is, $w_k(e_s) = p_k(e_s)$ for all s. If $h < 0$, there are different possibilities, according to which education levels are employed and which are not. In our setup,[5] employed workers will have (weakly) higher education levels than those of the unemployed. At any equilibrium, profits derived from any match with a worker, independently of the education level, will be the same, and will equal the profit that would follow from hiring an unemployed worker, whose outside option is z. Let $\hat{s}$ be such that, in the market with $h \leq 0$,

$$\sum_{\hat{s}+1}^{S} i_{e_s} < j \leq \sum_{\hat{s}}^{S} i_{e_s}. \tag{1}$$

In words, $e_{\hat{s}}$ is the lowest education level such that a worker is employed. Then, for $s \geq \hat{s}$, we will have

$$w_k(e_s) = p_k(e_s) - p_k(e_{\hat{s}}) + z,$$

so that a worker with education $e_{\hat{s}}$ is indifferent between being employed and unemployed, and firms earn the same profits from hiring any worker, including those that are unemployed. Summarizing, wages will be

$$w_k(e_s) = \begin{cases} p_k(e_s) & \text{if } h > 0 \\ p_k(e_s) - p_k(e_{\hat{s}}) + z & \text{if } h \leq 0 \text{ and } s \geq \hat{s} \\ z & \text{if } h \leq 0 \text{ and } s < \hat{s} \end{cases} \tag{2}$$

where $\hat{s}$ is given by (1).

Correspondingly, a firm earns, in each job, a profit given by

$$\pi_k(\mathbf{i}, j) = \begin{cases} 0 & \text{if } h > 0 \\ p_k(e_{\hat{s}}) - z & \text{if } h \leq 0 \end{cases},$$

where, again, $\hat{s}$ is given by (1).

Given wages, profits, and job and worker flows across markets, we can characterize the value functions for a new job and for a worker given his level of education.

The value of a new job in a type$-k$ market, $J_k(N_k)$, follows from

$$r J_k(N_k) = E\left[\pi_k(\mathbf{i}, j)\right] - l J_k(N_k),$$

where $\pi_k(\mathbf{i}, j)$ is the flow profit of a new job in a type$-k$ market, as described above.

Creating a new job costs $\kappa > 0$, so that the following no-profit condition should hold:

$$J_k(N_k) = \frac{E\left[\pi_k(\mathbf{i}, j)\right]}{r + l} = \kappa. \tag{3}$$

The value of a new job in a type$-k$ market weakly falls with N_k. As N_k grows, the proportion of markets where there is unemployment falls. In those markets

---

[5] This will be the case in the signaling equilibria we will examine.

8

where a single education level is present, that means that wages may go up from z to the value of a worker's expected productivity. In markets where several education levels are present, raising N_k implies that workers with (weakly) lower education levels are hired, so, again, π_k falls.

Let us now turn to workers. For any worker, the value of being in a type$-k$ market with an education level e is characterized by

$$
\begin{aligned}
r W_e^k \left(\mathbf{i}, j \right) & = w_k \left(e \right) + q \left[\overline{W}_e - W_e^k \left(\mathbf{i}, j \right) \right] \\
& + \max \left\{ E_{\mathbf{i'}, j'} \left[W_e^k \left(\mathbf{i'}, j' \right) \right], \overline{W}_e - \Gamma \right\} - W_e^k \left(\mathbf{i}, j \right),
\end{aligned}
$$

where $\overline{W}_e$ is the expected value of being allocated at random to any market given education level e. The third term on the right-hand side of this expression follows from the fact that the worker may voluntarily be reallocated: whenever $\overline{W}_e - \Gamma > E_{\mathbf{i'}, j'} \left[W_e^k(\mathbf{i'}, j') \right]$, the worker would decide to switch markets (i.e., $\varphi_e^k(\mathbf{i}, j) = 1$). Note that this expectation is taken over the process of $\mathbf{i}$ and j, which, as we mentioned, change according to the multinomial distributions described above.

The expected value of being reallocated at random to a new market, $\overline{W}_e$, follows from taking expectations over the joint distribution of $(\mathbf{i}, j)$. Specifically,

$$
\overline{W}_e = E \left[W_e^k(\mathbf{i}, j) \right],
$$

where the expectation is taken over all different market types. More concretely, this expectation is taken according to the cdf

$$
G(\mathbf{i}, j) = \sum\nolimits_{k=1}^{K} L_k G_k(\mathbf{i}, j),
$$

where each G_k is a cdf in markets of type k.

Finally, note that for any values state variables may take, the value of being in a market cannot fall below $\overline{W}_e - \Gamma$. Then, we asume that if a worker is reallocated to a market where the number of workers is excessive, or where the returns to his abilities are low, he would switch markets again instantly.

2.4   Signaling

We have described how the economy operates given workers' educational choices. We turn now to the signaling role of education. Our approach is based on a very simple version of Spence (1973).

As we mentioned above, before being randomly assigned to his first market, each worker selects an education level $e \geq 0$. This choice will be the only worker characteristic that any firm will observe later on in the labor market. The worker's utility is

$$
u(W, e, \psi, \varepsilon) = W - c(e, \psi, \varepsilon),
$$

where W is the value of his income in future employment, and the cost of education is

$$
c(e, \psi, \varepsilon) = e \psi^{-a} \left[t + \exp(\varepsilon) \right],
$$

9

with $a > 0$ and ε being an i.i.d. random shock to costs distributed according to $F(\varepsilon)$. Then, the cost (and the marginal cost) of education decreases with the worker's ability.

From our previous description of how labor markets work given educational choices, we will have

$$W = \overline{W}_e = E\left[W_e^k(\mathbf{i}, j)\right],$$

where the expectation is taken according to the stationary distribution of $\mathbf{i}$ and j, as explained above.

Clearly, then, workers will select an education level e^* such that

$$\arg\max_e \left\{ \overline{W}_e - e\psi^{-a}\left[t + \exp(\varepsilon)\right] \right\}. \tag{4}$$

The noise present in the cost function for education will be our instrument to evaluate how well education may play a signaling role in this economy. A worker with a high value of ψ may choose a low level of education -even if choosing a higher level of e is rewarded by the labor market- if ε is sufficiently high. More dispersion in $F(\varepsilon)$ may then make education perform worse as a signaling device.

2.5 Equilibrium

Combining the signaling and job matching aspects of the model, an equilibrium in this economy will be given by (i) a rule $e^*(\psi, \varepsilon)$ that specifies workers' educational choices, (ii) for each market type k, a wage schedule $w_k^*(e)$, (iii) for each market type k, a number of jobs per market N_k, (iv) for each market type k and education level e, a quitting rule $\varphi_e^k(\mathbf{i}, j)$, and (v) for each market type k, a cdf $G_k(\mathbf{i}, j)$ such that

- $e^*(\psi, \varepsilon)$ solves (4) for all (ψ, ε), given $E\left[W_e^k(\mathbf{i}, j)\right]$, computed according to $G(\mathbf{i}, j) = \sum_{k=1}^K L_k G_k(\mathbf{i}, j)$,

and, for each market type k,

- N_k follows from (3),

- $w_k^*(e)$ follows from (2),

- for each education level e, $\varphi_e^k(\mathbf{i}, j)$ equals one (zero) if $\overline{W}_e - \Gamma > E_{\mathbf{i}', j'}\left[W_e^k(\mathbf{i}', j')\right]$, where the expectation is computed according to $G_k(\mathbf{i}, j)$, and

- $G_k(\mathbf{i}, j)$ satisfies the worker reallocation and job creation conditions described above.

3   A simple case

So as to describe in a clearer way the effect that signal quality may have, let us resort to a simple version of the general model described above.

We will now concentrate on equilibria where educational choices generate partial separation at the lowest possible educational cost. Specifically, each worker type will be associated to an education level. That level will be the lowest one such that a worker of the type that is immediately worse will never select it, for any value of ε. Then, workers of a given type will choose the corresponding education level for low values of ε. If ε grows, however, that choice will eventually be too costly, and the worker will rather select the educational level associated to a lower type. Partial separation will follow: the highest level of education will only be chosen by (some) workers of the highest type, whereas each lower level will be selected by workers of the type associated to it and by some workers of better types, for whom education is costly enough.

To examine this class of equilibria in our model, suppose $\psi \in \Psi = \{\underline{\psi}, \overline{\psi}\}$. Furthermore, assume as well that the shock to the cost of education, ε, is normally distributed, with mean zero and variance σ_ε^2.

In the equilibria we will examine, then, two education levels will be selected with positive probability: $\underline{e}$, associated with $\psi = \underline{\psi}$, and $\overline{e}$, associated with $\psi = \overline{\psi}$, with $\overline{e} > \underline{e}$. We take strategies such that all workers of type $\underline{\psi}$ select $\underline{e}$, for any value of ε. A worker of type $\overline{\psi}$ chooses $\overline{e}$ unless his cost shock ε takes a very large value. All intermediate education levels are not rewarded by the labor market: we assume that market beliefs about worker productivity for those education levels coincide with those that would follow after observing $\underline{e}$.[6]

Naturally, the corresponding incentive condition has to hold for low types: for all ε, they have to prefer $e = \overline{e}$ to $e = \underline{e}$. That is,

$$E\left[W_{\underline{e}}^k(\mathbf{i}, j)\right] - \underline{e}\underline{\psi}^{-a}[t + \exp(\varepsilon)] \geq E\left[W_{\overline{e}}^k(\mathbf{i}, j)\right] - \overline{e}\underline{\psi}^{-a}[t + \exp(\varepsilon)]$$

for all ε.

So that (partial) separation follows at the lowest possible cost, we select the equilibrium where education takes the lowest possible values. Then, $\underline{e} = 0$ and $\overline{e}$ is, given parameter values, the lowest value of e such that all low-type workers weakly prefer $\underline{e}$. That is, $\overline{e}$ is such that

$$E\left[W_{\underline{e}}^k(\mathbf{i}, j)\right] = E\left[W_{\overline{e}}^k(\mathbf{i}, j)\right] - \overline{e}\underline{\psi}^{-a}t,$$

as Figure 1 shows. The figure depicts indifference curves for a $\underline{\psi}$-type worker when $\varepsilon \to -\infty$. Clearly, given the expected labor income premium associated with education $\overline{e}$, in that limiting case the worker is indifferent between choosing that education level and not educating at all. Thus, all low-type workers prefer $e = 0$, independently of their cost shocks.

---

[6] Since those education levels are chosen with zero probability, firms' beliefs after observing them are arbitrary.

Figure 1: Signaling model: indifference curves of low skilled
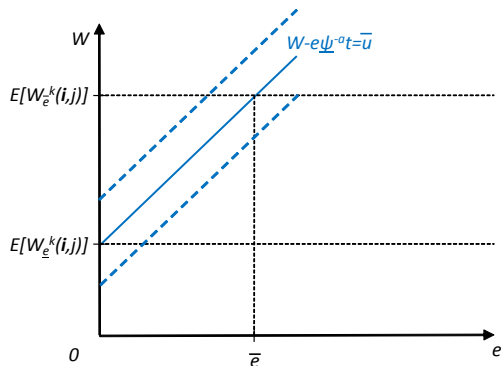
In the case of high types, Figure 2 below depicts their situation. As ε grows, a high type's indifference curves become steeper, since the marginal cost of education grows. The figure shows indifference curves for high types that select $e = \bar{e}$, given the expected labor income premium, for different values of ε.

Figure 2: Signaling model: idiosyncratic costs and partial pooling
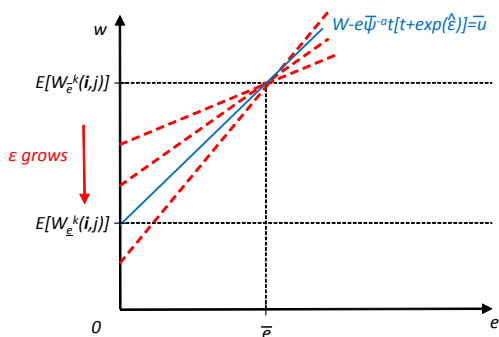
Eventually, if ε grows enough, a high type will prefer mimicking a low type by choosing $e = 0$ rather than $e = \bar{e}$. Hence, there will be a cutoff value for ε, $\widehat{\varepsilon}$, given by

$$ E\left[W_{\underline{e}}^k(\mathbf{i}, j)\right] - \underline{e}\overline{\psi}^{-a}\left[t + \exp(\widehat{\varepsilon})\right] = E\left[W_{\bar{e}}^k(\mathbf{i}, j)\right] - \overline{e}\overline{\psi}^{-a}\left[t + \exp(\widehat{\varepsilon})\right]. $$

Then, high-type workers with $\varepsilon < \widehat{\varepsilon}$ ($\varepsilon > \widehat{\varepsilon}$) would choose $\overline{e}$ (respectively, $\underline{e} = 0$).

Notice that $\widehat{\varepsilon}$ depends on expected labor income, and parameters such as a and t, but does not depend on σ_ε. Let the parameters be such that when $\sigma_\varepsilon \to 0$ a separation equilibrium in the signaling problem can arise. In such a case, and fixing expected labor incomes and $\widehat{\varepsilon}$, a higher variance for ε implies that more high-type workers select a lower education level. In essence, a higher variance for ε reduces education's effectiveness as a signal. In these equlibria, a higher education level still reveals high productivity, but is less frequently chosen. A lower (in our case, zero) education level becomes less informative, as it is chosen by all low types and a larger portion of high types. Fewer workers tend to separate using education as a signal.

Needless to say, though, this is not the end of the process. As signal quality becomes worse, and educational choices change, both $E\left[W_{\underline{e}}^k(\mathbf{i}, j)\right]$ and $E\left[W_{\bar{e}}^k(\mathbf{i}, j)\right]$ change as well. Indeed, as the signal is less informative, the labor income premium will fall, since workers will more often be stuck at an inefficient labor match, and -given our equilibrium selection- so will $\bar{e}$, the education level selected by those high types that do separate. In addition, the cutoff value $\widehat{\varepsilon}$ will also vary. As our quantitative exercises below will show, overall labor market performance will suffer.

4   Data and empirical findings

In this section we describe the empirical findings that support the idea that education could be considered a noisy signal of skills. Additionally, these empirical findings will allow us to calibrate our model. First, we describe the data. Then, we discuss the evidence of higher mismatch and worse educational quality in Latin America, as compared to OECD countries. Finally, we present the results that follow from a regression analysis on wages, which are relevant to further identify signaling problems and for the calibration of our model.

4.1   Data

Our first source of information is the Survey of Adult Skills, the main output of the Programme for International Assessment of Adult Competencies (PIAAC), run by the OECD. This survey is designed to be representative of the population between ages 16 and 65 in very diverse countries using a harmonized methodology. It includes data on occupation, education and skills (numeracy, literacy and problem solving). Notably, though, the PIAAC provides this information for OECD countries. There is no Latin American country with available data in the program so far -results on Chile will be available in the future.

As for Latin America, there are a few important sources that provide information on skills (both cognitive and noncognitive) for adults in the region. Our main source is the *Encuesta Nacional de Habilidades y Mercado Laboral* (EN-HAB), carried out by the World Bank in Peru in 2010 (see Yamada et al., 2014). In this survey, cognitive skills were measured through several tests on numeracy,

problem solving, working memory, verbal fluency and language. Noncognitive variables are based on the GRIT scale and the Big Five personality traits.

Additionally, we use data from the World Bank's Skills Measurement Program, STEP. This source covers several developing countries between 2012 and 2014, including Bolivia and Colombia. It is a household survey of adults (between 15 to 64 years of age) with the objective of analyzing workers' skills, as well as their formal education and labor market outcomes. It recovers both cognitive (comparable to the PIAAC information) and socio-emotional skills (Big Five, hostile attribution bias and other variables related to decision making).

Our main objective when using these data sources will be to explore differences between Latin American and OECD countries. We first focus on the allocation of skills across occupations and we then present the relevant wage regressions.

4.2   Mismatch of skills

A growing empirical literature examines mismatch from an empirical point of view. Mismatch is usually measured by comparing the skills required for each job/occupation, according to a standard job description, with the skills that workers employed in each occupation have. The latter could be identified by using information on educational levels, as in Sicherman (1991) or Chevallier (2003). To follow this approach, there are many available information sources, both household surveys and census data.

Alternatively, skills required in a job could be compared with actual worker abilities. This is the path taken, for example, in Pellizzari and Fichen (2013), Allen and van der Velden (2001) and Desjardins and Rubenson (2011), among others. Needless to say, the main hurdle in this approach is how to measure worker skills directly.

We draw from these examples to measure mismatch as in Pellizzari and Fichen (2013). That paper uses PIAAC information on skills and job descriptions to explore whether workers are under- or over-skilled for their current jobs. Well-matched workers are identified using two survey questions: (i) "*Do you feel that you have the skills to cope with more demanding duties than those that are required to perform in your current job?*", and (ii) "*Do you feel that you need further training in order to cope well with your present duties?*". If an individual gives a negative answer to both questions, he is considered well matched according to this method. Then, minimum and maximum skills in the group of well-matched workers determine the skill thresholds by occupation that are used to identify those who are under- or over-skilled. For example, over-skilled workers would be those with skills that are higher than the maximum among those who have been identified as well-matched.

We follow the same procedure using PIAAC information at the two-digit occupation level, and find that 93% (92%) of workers can be considered well-matched according to their numeracy (respectively, literacy) skills in Germany (see Table 1). We can exploit as well standardized numeracy and literacy scores in ENHAB. When applying the same thresholds for Peru, we find that 78%

Table 1: Distribution of skills by occupations (PIACC & EHAB)

		Germany	Peru
(A) Mismatch			
Numeracy			
	Well-matched	0.93	0.78
	Under-skilled	0.02	0.03
	Over-skilled	0.05	0.19
Literacy			
	Well-matched	0.92	0.87
	Under-skilled	0.02	0.02
	Over-skilled	0.06	0.11
(B) Variance			
Numeracy			
	Between	0.18	0.14
	Within	0.82	0.86
Literacy			
	Between	0.16	0.10
	Within	0.84	0.90

Notes: Panel (A) reproduces the methods presented in Pellizzari and Fichen (2013) to identify under- and over-skilled workers. Panel(B) reports the proportion of variance within and between groups using variance decomposition methods.
Source: PIAAC and ENHAB.

(87%) of workers can be identified as well-matched considering numeracy (respectively, literacy). Notice that our underlying assumption in this exercise is that, without distortions, jobs would require the same relative skills in Peru as in Germany. By standardizing our measures of cognitive skills in different countries, we are assuming that the range of skills is similar in both.

Additionally, we can analyze the distribution of skills across occupations. In particular, we decompose the total variance of standardized scores in skills, using a within- and between- group variance decomposition. The intuition behind this exercise is that a higher within-group variance proportion is associated to a higher likelihood that skills are mismatched. We find that the within variance proportion is approximately five points higher in Peru than in Germany. As an example, numeracy skills variance within occupations is 82% of total variance in Germany, while it is 86% in Peru. Results are similar for literacy skills.

These results suggest that labor markets in Peru exhibit more mismatch than those in Germany.

4.3   Skills distribution and signaling

There are different pieces of evidence that suggest that education does not perfectly signal worker's abilities in Latin American countries. In order to discuss this point, we will first resort to international skills assessments that build scores of cognitive skills in a standardized way between countries. We use PIAAC results for OECD countries and STEP data for Latin American countries.[7]

The main quality indicator for education that we want to emphasize is the distribution of skills within each level of education. We then associate higher heterogeneity in abilities within each level of education with a less precise signal.

Table 2 displays the standard deviation of literacy scores. It shows that skills in Latin American countries are more heterogeneous given education. In particular, not only is the standard deviation of literacy scores larger overall, but it is also larger within each level of education. This difference is more noticeable for primary, high school and tertiary education. On the other hand, panel (B) of table 2 shows the same indicator, but after controlling for country fixed effects by simple OLS regressions (as the ones presented in tables 3 and 4). Differences between OECD and Latin American countries tend to be smaller but are still strong.

The results of an OLS regression of literacy scores on indicator variables for each level of education -where primary is the base category- are summarized in Table 3. All coefficients are measured with higher variance in Latin American countries than in the OECD. Standard errors (in parenthesis in the table) are about three times larger than in OECD countries, even after adding country fixed effects (columns 3 and 4).

As described above, we use as well Peruvian data, from ENHAB. The methodology of this source, however, is not necessarily comparable to the mea-

---

[7]PIAAC covers Austria, Belgium, Canada, Denmark, Estonia, Finland, France, Great Britain, Germany, Ireland, Italy, Japan, S.Korea, Netherlands, Norway, Poland, Russia, Slovakia, Spain, Sweden and the US. STEP countries are Bolivia and Colombia.

Table 2: Standard deviation of literacy score

Education	(A) Raw scores		(B) Residuals of regressions	
	OECD	Latin America	OECD	Latin America
Primary	49.10	77.72	48.42	70.63
High School	39.99	52.77	38.92	48.94
Tertiary	39.84	58.54	39.08	58.54
Some College	38.04	47.64	36.43	43.64
College	39.75	39.46	37.84	39.24
Total	48.47	70.80	40.07	56.61

Notes: Latin American countries are Colombia and Bolivia. Sample is restricted to individuals from 25 to 49 years of age and with comparable levels of education between countries according to ISCED.
Source: PIAAC and STEP literacy scores for selected countries.

sures in PIAAC and STEP. Accordingly, we standardize literacy scores by country (i.e. we substract the mean and divide by the standard deviation). Given that the point here is to identify variance within each education level rather than total variance, we believe that standardized variables are informative.
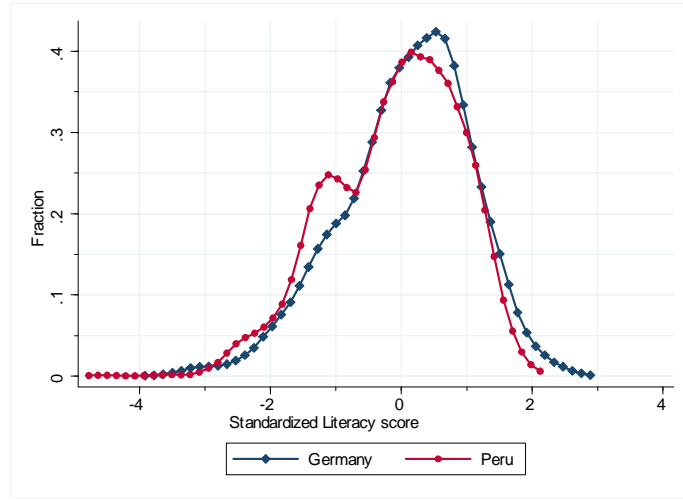
Figure 3 depicts the density of this standardized measure of literacy skills for Peru. The analogous distribution for Germany is also included for the sake of comparison. In Peru, the skills distribution is bimodal, with a relevant group of individuals with lower skills than the median. While this type of bimodal distribution is also found in Colombia and Bolivia, the corresponding distributions for all OECD countries, from PIAAC, are unimodal. Importantly, this bimodality is still present for the subsample of those with secondary and tertiary education. This bimodal distributions by education level suggest that the higher skills variance we have found in Latin American countries may follow from a long left tail, from the lowest part of the skills distribution.

The last two columns in table 3 present the coefficients that obtain when regressing the standardized measure of literacy on schooling level and country fixed effects. Even after standardizing, we find more dispersion in coefficients, with standard errors approximately doubling those for OECD countries. Table 4 shows that these results are robust when the same sample is restricted to males.

Another relevant and noticeable point in the regressions in table 3 is that, while all levels of education are significantly different from primary education, each level does not always imply a significant improvement when compared to the previous one. As an example, tertiary education does not imply a significant improvement when compared with high school in Latin American countries (see columns 2 and 4). Something similar happens when we analyze ENHAB data for Peru.

To provide clearer evidence on this issue, we regress standardized scores on a

Figure 3: Distribution of standardized literacy

series of indicator variables that identify all observations with education higher than a given level. In this sense, all coefficients should be read as increments from previous level. Table 5 compares OECD countries to Peru using this definition of education. First, we still observe higher standard errors in Peru for each coefficient. Additionally, we find that "some college" is not correlated with significantly higher cognitive skills in Peru when compared to "tertiary," and that "college" is not correlated with higher literacy scores when compared to "some college." Importantly, this is not the case in OECD countries. This is another type of evidence of relatively bad quality of education, since these levels of education do not signal differential skills.

Table 3: Regressions of literacy scores on education

Variable	(1) Literacy	(2) Literacy	(3) Literacy	(4) Literacy	(5) Lit. Stand.	(6) Lit. Stand.
Countries	OECD	L.A.	OECD	L.A.	OECD	L.A.
High School	41.81***	64.04***	38.79***	56.28***	0.82***	0.82***
	(0.81)	(5.21)	(0.85)	(4.47)	(0.02)	(0.05)
Tertiary	48.33***	50.56***	47.54***	77.99***	1.03***	1.21***
	(1.73)	(7.69)	(1.77)	(7.95)	(0.04)	(0.07)
Some College	55.60***	76.94***	53.85***	76.51***	1.15***	1.27***
	(0.90)	(5.73)	(0.93)	(4.86)	(0.02)	(0.06)
College	71.22***	96.92***	69.17***	103.34***	1.48***	1.60***
	(0.79)	(5.07)	(0.82)	(4.65)	(0.02)	(0.05)
Constant	234.96***	167.47***	256.71***	140.04***	-0.97***	-0.85***
	(0.69)	(4.25)	(1.46)	(4.69)	(0.03)	(0.05)
Country dummies	No	No	Yes	Yes	Yes	Yes
R^2	0.27	0.26	0.32	0.36	0.25	0.33
N	42,607	2,614	42,607	2,614	42,607	4,002

Notes: OLS estimations (standard errors) of the literacy scores on indicator variables of education. In columns (3)-(6) we add fixed effects by country. In columns (1) to (4) we use PIAAC and STEP comparable literacy measures. In columns (5) and (6) we use standardized measures of literacy skills and we include ENHAB measure of literacy for Peru. Sample is restricted to individuals from 25 to 49 years of age and with comparable levels of education between countries according to ISCED. *, **, and *** indicate significance at the 10%, 5% and 1% levels.
Source: PIAAC, STEP and ENHAB literacy scores.

Table 4: Regressions of literacy scores on education - Males

Variable	(1) Literacy	(2) Literacy	(3) Literacy	(4) Literacy	(5) Lit. Stand.	(6) Lit. Stand.
Countries	OECD	L.A.	OECD	L.A.	OECD	L.A.
High School	42.51***	58.53***	40.29***	51.29***	0.85***	0.76***
	(1.21)	(7.78)	(1.26)	(6.33)	(0.03)	(0.09)
Tertiary	43.80***	31.89**	42.82***	62.11***	0.92***	1.17***
	(3.41)	(14.37)	(3.54)	(14.74)	(0.07)	(0.13)
Some College	53.92***	71.48***	55.26***	75.47***	1.17***	1.29***
	(1.43)	(8.17)	(1.43)	(6.93)	(0.03)	(0.08)
College	72.14***	90.13***	70.16***	102.92***	1.50***	1.56***
	(1.17)	(7.44)	(1.21)	(7.13)	(0.03)	(0.08)
Constant	236.93***	177.34***	255.19***	147.12***	-0.99***	-0.83***
	(0.99)	(6.22)	(2.47)	(7.02)	(0.05)	(0.08)
Country dummies	No	No	Yes	Yes	Yes	Yes
R^2	0.28	0.24	0.33	0.35	0.26	0.32
N	18,484	1,079	18,484	1,079	18,484	1,565

Notes: *, **, and *** indicate significance at the 10%, 5% and 1% levels. See notes from Table 3
Source: PIAAC, STEP and ENHAB literacy scores.

Table 5: Regressions of standardized literacy scores on education levels - Comparing OECD with Peru

	(1)	(2)	(3)	(4)
Countries	OECD	Peru	OECD	Peru
Sample	All	All	Males	Males
Primary	-0.88***	-0.92***	-0.99***	-0.78***
	(0.03)	(0.07)	(0.05)	(0.14)
High School	0.82***	0.82***	0.85***	0.77***
	(0.02)	(0.09)	(0.03)	(0.17)
Tertiary	0.22***	0.54***	0.07	0.69***
	(0.04)	(0.08)	(0.07)	(0.14)
Some College	0.11***	0.32***	0.25***	0.17
	(0.04)	(0.09)	(0.07)	(0.12)
College	0.34***	0.11	0.33***	0.02
	(0.02)	(0.09)	(0.03)	(0.11)
R^2	0.26	0.39	0.27	0.42
N	42,607	1,388	18,484	486

Notes: OLS estimations (standard errors) of standardized literacy scores on indicator variables of education and fixed effects by country. Education is a set of dummy variables that identify that the individual has completed the level. As an example, High school is "at least high school education", etc. Then, each coefficient should be read as the addition of the level to the previous one. Sample is restricted to individuals from 25 to 49 years of age and with comparable levels of education between countries according to ISCED. In columns (3) and (4) we restrict further the sample to males. *, **, and *** indicate significance at the 10%, 5% and 1% levels.
Source: PIAAC and ENHAB standardized literacy scores. Literacy scores are standardized by country.

So far, we have concentrated on measures of cognitive skills. In particular, we have shown that some cognitive skills are correlated with schooling level, but that skills seem to be more dispersed in Latin American countries within each level of education. We interpret this as evidence of the fact that education provides a worse signal.

A worker is not only productive because of cognitive skills, but also because of noncognitive skills, as the literature has shown. If both types of skills are important, then, the quality of education should be measured by its performance in building and signaling both of them.

To examine this issue we use information from ENHAB, Peru. This source is rich in the diversity of cognitive and noncognitive measures it provides, and thus allows us to consider both types of skills. In order to provide simpler measures, we use exploratory factor analysis to build single factors that summarize several variables in a single combination. In particular, we estimate a single factor of cognitive skills that captures information from scores in literacy, numeracy, verbal fluency and working memory. Additionally we estimate a single factor of noncognitive skills, combining the scores of six measures of socioemotional abilities (Big Five and Grit).

Finally, we combine all these cognitive and noncognitive skills in one single measure. For that purpose, and following the idea behind latent variables (see Heckman et al., 2006) we included all these variables and two variables of income (the log of hourly wages and the log of monthly labor income) to extract a single factor combining them all. Intuitively, if a single latent variable characterized an individual's skills that influence productivity, this latent variable should be related to scores in standard tests of cognitive and noncognitive skills and, at the same time, to the income the individual receives. To obtain this single indicator, we run exploratory factor analysis on all these variables on the sample of wage earners and then, taking the factor loadings from this sample, we predict factor levels for the whole sample.

Table 6 shows the standard deviation of these factors (cognitive, noncognitive and a single factor for all skills) by education. There is no important difference in these factors' overall standard deviations, but they have within variance differences when analyzing groups by education. The standard deviation is larger for the noncognitive factor. Furthermore, the standard deviation for the single factor of skills is higher than that for cognitive skills. This suggests that education performs worse as a signal of noncognitive skills than as one of cognitive skills.

The cumulative distributions of cognitive and skills factors for some education groups are displayed in Figure 4. It is apparent that these skills do not necessarily vary by much when the education level changes. In particular, when we compare "some college" with "college," skills are not distributed differently between the two groups. Not only are they similar in mean, but also in terms of the whole distribution.

To sum up, we have provided additional evidence suggesting that the quality of education as a means to signal cognitive ability is poorer in Latin American than in OECD countries. We have documented as well that education has even

Table 6: Standard deviation of one factor measures of cognitive, noncognitive and overall skills - Peru

Education	Cognitive	Noncognitive	Skills
Incomplete primary	0.64	0.92	0.83
Complete primary	0.61	0.95	0.85
Incomplete High School	0.67	0.86	0.76
Complete High School	0.71	0.85	0.76
Incomplete Tertiary	0.60	0.97	0.83
Complete Tertiary	0.67	0.85	0.76
Incomplete College	0.61	0.87	0.77
Complete College	0.63	0.82	0.70
Postgraduate	0.54	0.90	0.85
Total	0.86	0.90	0.86

Notes: Standard deviations of factors measures by education level. All the factors were generated using exploratory factor analysis. Cognitive measure was generated using five measures of cognitive abilities (literacy, working memory, numeracy and two measures of verbal fluency). Noncognitive measure was generated using six measures of socioemotional abilities (Big Five and Grit). The one factor measure for overall skills was generated using all previous measures and two measures of labor income (log of hourly wages and log of monthly labor income).
Source: ENHAB.

higher problems to signal noncognitive abilities in Peru.

4.4   Education, skills and allocation in occupations

Let us now provide additional evidence of allocation within occupations. For that purpose, we provide a very broad definition of occupation groups, education levels and skills that will be useful for the calibration exercise we will later present. To do so, we resort to ENHAB information from Peru, and we concentrate on individuals aged between 23 to 40.
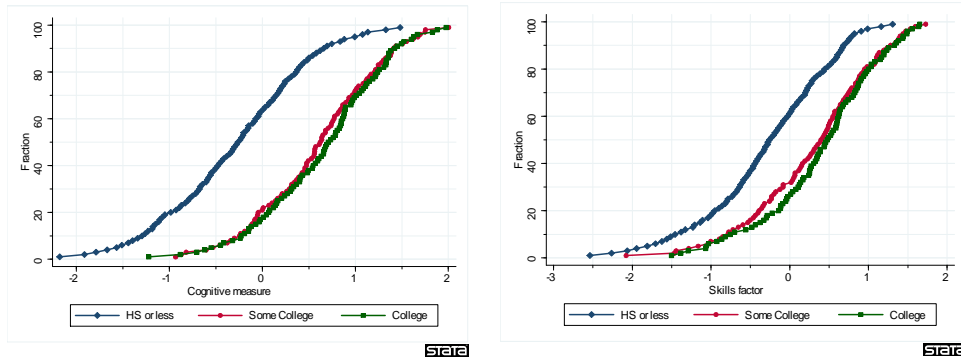
We define an unique indicator of skills by running a factor model on all variables regarding skills, cognitive and noncognitive, and selecting only one factor. We defined as high skilled those observations above the 70th percentile in the distribution of this skills factor.

Additionally, we divide individuals according to their education into two groups. We consider "highly educated" those workers with secondary or higher education, while we regard the remaining workers as having "low education."

Table 7 shows the proportion of individuals in each group according to their education and skills. It is clear that, while high education and high skills are correlated, this correlation is not perfect, as about 30% of the low skilled have high education.

We now turn to the allocation of these workers in different occupations. For that purpose we broadly divide occupations into three categories: (1) managers and profesionals; (2) technitians; and (3) non qualified.

Figure 4: Cumulative distribution of cognitive and noncognitive skills by education - Peru

Table 8 reports the distribution of workers according to their education and skills in these three categories. We first find that occupations differ widely according to education: 97% of workers in occupation 1 have high education, while 78.4% do in occupation 2 and 23% in occupation 3. However, skills are not so sharply allocated among occupation categories: 45%, 55% and 24% respectively. The difference in allocation according to education and according to skills is possibly related to education performing poorly as a signal.

4.5 Wage regressions

As we described in Section 1, signaling theory has been empirically tested by a large literature. However, this literature has not used information of actual skills. In our context, this information could be crucial to determine whether education has a signaling effect through wage regressions. If education is a perfect device for signaling skills, then returns to education should be the same as returns to skills. In addition, if education was not a signaling device but only a way of acquiring (observable) skills, then all returns to education could be measured by the increase in skills between educational levels, and education would have no effect after controlling for skills.

On the contrary, if education was an imperfect signal of skills, these returns would differ. Moreover, even comparing workers with the same skills, it is pos-

Table 7: Education and skills - ENHAB - Peru

	Low skills	High skills
Low education	0.494	0.128
High education	0.206	0.171

Table 8: Education and skills by occupation - ENHAB - Peru

	1. Managerial	2. Technicians	3. Other
	1. Managerial	2. Technicians	3. Other
Prop. of highly educated	0.97	0.78	0.23
Prop. of high skilled	0.44	0.55	0.24

sible for education to have a positive impact on wages if signaling is imperfect. This would happen if there was variation in education levels among workers with the same skills, and if firms' expectations are consistent with a positively sloped wage schedule with respect to education.

Thus, taking our model as a basis, we analyze the relationship between wages, skills and education so as to test for signaling quality: returns to education, after controlling for skills, will imply imperfect signaling.[8]

To do so, we run regressions of the log of labor income on indicators of skills, education and other controls, including sex, age, age squared, as well as indicator variables regarding household member and marital status over our sample. Indicators of skills and education are defined as in the previous analysis.

Table 9 summarizes the results of these regressions. The first column shows that employed workers with high education earn .52 log points more than those with low education. Wages are also related to skills: workers with high skills earn .31 log points more than the low-skilled, as shown in the second column. Both of these coefficients are highly significant.

We compute the residuals of this regression to analyze whether they can be explained by education. The third column shows the results of a regression of these residuals on education. The coefficient implies that higher education yields .39 log points higher wages, even after controlling for skills.

An alternative way of showing that education is still relevant even after considering skills is by a joint regression of wages on both skills and education.

---

[8] We postpone the discussion of measurement error in skill variables that could affect this implication.

Table 9: Regression of log wages on skills and education indicators - ENHAB - Peru

	(1)	(2)	(3)	(4)	(5)
	Education	Skills	Residuals	Edu & Skills	By occupation
High education	0.52***		0.39***	0.47***	
	(0.11)		(0.10)	(0.12)	
High skills		0.32***		0.19*	.18
		(0.11)		(0.12)	(.20)
Occup. 2					-0.35*
					(0.19)
Occup. 3					-0.75***
					(0.16)
Occup. 2 x High Skills					-0.02
					(0.29)
Occup. 3 x High Skills					0.13
					(0.25)
R^2	0.126	0.07	0.06	0.14	0.17
N	362	362	362	362	361

The table's fourth column shows that highly educated workers receive .47 log points higher wages, while the wages of those with high skills are only .19 log points higher.

On the whole, we find that wages are correlated not only with skill indicators, but also with education levels. In our model, this is only consistent with education being a bad signal of skills.

The last column of Table 9 shows the results that follow from regressing wages on a skills dummy interacted with occupation dummies (identifying the three broad occupation categories described above). We will use these results when calibrating our model.

Before turning to quantitative results, it is important to emphasize that our main conclusions rest on the assumption that cognitive and noncognitive indicators are a relevant and accurate measure of skills valued in the labor market. If these variables were completely noisy, then our method would be misleading. For that reason, it is important to report that skills are significant explanatory variables of wages, which implies that these variables are not pure noise. Additionally, we find similar qualitative results when we use different measures of skills and education, including factors for skills and several levels of education. In all cases, we find that education is still relevant even after

controlling for skills.

5   Calibration

Implications of the model   According to our model, evidence of mismatch may be found in the distribution of education among occupations and in unemployment. Indeed, without labor market frictions all workers would be employed and would be allocated to the occupation where their productivity is highest. Additionally, with no frictions, it is plausible that some markets or occupations would close if their productivity is lower than that in some other market for all skill levels. In the model, mismatch falls when the number of workers per market increases, as well as when mobility costs vanish. In such a case, there is no reallocation problem and no uncertainty, so firms create as many jobs as there are available workers.

An efficient allocation would arise in a context in which there are no frictions (no reallocation costs and no job creation costs) and no signaling problems (skills are observable by firms). Jobs would open up to satisfy a zero-profit condition, so that all wages would equal workers' productivity. Workers would reallocate to those market types where productivities are higher. In that case, market types with lower productivities for all types of skills would close, emphasizing that these occupations subsist just because of labor market frictions.

In the model, poor signaling influences the variability of skills among workers in the same education level, with more skill dispersion indicating poorer signaling. Additionally, similarly skilled workers can be paid differently, according to the signal rather than to their productivity. In other words, when signaling is noisy skills and education are not mapped one to one, and even after controlling for skills, education is positively related to wages, the probability of being employed and income. We make use of these two observations when linking the data to our model.

Parameters   We set the unit of time equal to a quarter and the interest rate r to 1%. Abilities take discrete values, $\psi = \{1, 2\}$, defining two types of workers. We set $M = 10$ and the proportion of workers with $\psi = 1$ as 0.5, so that we can interpret the results as those above or below the median in the distribution of skills. We consider three groups of occupations/markets, with proportions 0.1, 0.1 and 0.8.

For the signaling problem we assume that the distribution of idiosyncratic shocks are normal with mean zero, so that we reduce its parametrization to a single parameter: σ_ε, the idiosyncratic cost's standard deviation. We also asume $t = 0$ as another parameter of idiosyncratic distribution. We assume that education e takes two levels: 0 and $\bar{e}$, this last to be calibrated in equilibrium. Importantly, the quantitative model is different from the one in Section 2 in that we now allow both types of skills to pool in the two levels of education, so that there would be low and high skilled workers both in low education and in high education. We allow for this distribution in order to link the model

with the data (see Table 7). Importantly, to analyze the impact of changes in the signaling quality of education, in Section 6, we change the parameters of idiosyncratic costs (σ_ε and t) but fix $\bar{e}$.

The parameters to be calibrated in equilibrium are those from the labor market (κ, z), those related to transitions in the labor market (q, l, Γ), to productivity (such as the functions $y_k(\psi)$), to the signaling problem, (a), and the parameter of the distribution $F(\varepsilon)$, σ_ε.

We must emphasize, though, that not all of our parameters are identified in the model, but only up to proportional relations.

To match these parameters we use the series of moments that we compute for Peru from ENHAB and ENAHO.[9] Most of these moments were presented in Section 4, with the exception of transition rates, which are described in the Apendix. These moments are shown in the second column of Table 10.

The first four moments are those related to the distribution of workers by education and skills, which are analogous to those presented in Table 7. These moments are key to to calibrate the signaling problem.

As is standard when calibrating the mismatch model to identify the exogenous transition probabilities, we use the unemployment rate and transition rates, in and out of unemployment (finding and separation rates). These are the next moments in the table. We also add transitions in and out of occupations: the probability of a worker moving between occupation groups. Occupation groups are defined by the skills required the occupation: unskilled, technical and professional as in Section 4.

Some of the moments for calibration follow from regression analysis. The regression of wages on skills by occupation, presented in Table 9 in Section 4, is directly linked with the production function by occupation, whose parameters it helps identify.

The last set of moments are related to the allocation of workers in occupations. These are analogous to those presented in Table 8 of Section 4, and help identify the endogenous allocation decision by workers in each market type.

Table 11 lists the parameter values found when minimizing the distance between data moments and those in the model. These values show a clear order according to occupation: high-skilled workers are more valuable in managerial and professional occupations, while their productivity falls in types of jobs. Low-skilled workers, though, are more productive in technical occupations. Productivity (and wages) in the third occupation category is lower than in any of the other two. In an efficient allocation this category would disapear. In other words, a planner would allocate all skilled workers to professional occupations and all the unskilled to technical ones. Additionally, given these values, aggregate per worker productivity in the efficient allocation would be 28*0.5+16*0.5=22.

_____

[9] The National Household Survey (ENAHO for its acronym in Spanish) is a panel of hoseholds for which we can follow workers between years. Thus, we use ENAHO to construct transition rates, which we then we transform into quarterly rates. We describe this source and methods in the Appendix.

Table 10: Model and data moments

Proportion of workers	(1) Model	(2) Data	(3) Source
Low skills and low education	0.43	0.43	ENHAB
Low skills and high education	0.07	0.07	ENHAB
High skills and low education	0.37	0.37	ENHAB
High skills and high education	0.13	0.13	ENHAB
Unemployment rate	0.11	0.11	ENHAB
Finding rate	0.14	0.12	ENAHO
Separation rate	0.06	0.05	ENAHO
Change of occupation rate	0.02	0.03	ENAHO
Wage regression, occup 2	-0.02	-0.05	ENHAB
Wage regression, occup 3	-0.62	-0.61	ENHAB
Wage regression, high skilled	0.15	0.35	ENHAB
Wage regression, occup 2 x hs	-0.06	-0.42	ENHAB
Wage regression, occup 3 x hs	-0.14	-0.14	ENHAB
Prop. of workers, occup1, ls	0.08	0.05	ENHAB
Prop. of workers, occup1, hs	0.11	0.10	ENHAB
Prop. of workers, occup2, ls	0.08	0.03	ENHAB
Prop. of workers, occup2, hs	0.11	0.17	ENHAB
Prop. of workers, occup3, ls	0.33	0.35	ENHAB
Prop. of workers, occup3, hs	0.28	0.30	ENHAB

When comparing productivity in the market with that in home production, we find that the gains of being employed are higher for skilled workers.

The costs of reallocation (Γ) and of job creation (κ) are not high and correspond to average productivity at the efficient allocation and about twice this value, respectively.

On the other hand, the job destruction rate is relatively low (1%) as compared to the quitting probability (5%).

Finally, in the calibrated model, the standard deviation of the idiosyncratic cost of education, σ_ε, is substantial, 1.25.

Column (1) of table 10 includes the simulated moments using the calibrated model, and a comparison with data moments in column (2). In some dimensions, the model is able to reproduce exactly the moments in the data. This happens, for example, in the case of education decisions. The model adequately captures the transition and unemployment rates, too. However, it does not perform so well in reproducing the proportion of workers by skills in occupations. For instance, that proportion is very different in the data and in the simulated model for technical occupations.

Besides calibrated moments, the model is also able to capture an additional important aspect: the regression of wages on skills is significant and the residuals of this regression on education still generate a significant and high coefficient, around 0.83.

Table 11: Parameter values

Parameter	Value
Job creation cost; κ	50
Home production ; z	8
Productivity in occ. 1 and $\psi = 1$;	28
Productivity in occ. 1 and $\psi = 2$;	14
Productivity in occ. 2 and $\psi = 1$;	24
Productivity in occ. 2 and $\psi = 2$;	16
Productivity in occ. 3 and $\psi = 1$;	12
Productivity in occ. 3 and $\psi = 2$;	11
Job destruction rate; l	0.01
Workers' reallocation rate; q	0.05
Workers' reallocation cost; Γ	25
Std. of the idiosyncratic cost of education; σ_ϵ	1.25
Idiosyncratic cost parameter; t	0
Education level; e	15.62
Cost function parameter; a	-0.906

Notes: parameter values of the quantitative model.

Figure 5 shows the distribution of workers and jobs in all markets -that is, the simulated counterpart of $G(\mathbf{i}, j)$. There are two relevant aspects of this distribution. First, the highest peak corresponds to the allocation of workers and jobs in lower skilled occupations. Given that there are many of these markets and few workers in each of them, the distribution is strongly concentrated there. The low-density portion of this distribution corresponds to the remaining two occupation categories, which require higher skills.

Figure 5: Distribution of workers in all markets

The distribution of workers and slack in jobs, h, in market type 1 -i.e. in managerial and professional occupations- is depicted in Figure 6. The typical market of this type would have between 15 and 20 workers and more jobs than workers. Nevertheless, the distribution is quite dispersed.

Figure 7 shows the joint distribution of workers and jobs in the case of unskilled occupations. Significantly, there are no highly educated workers in this distribution, but only workers without education. Given that we are analyzing (partial) pooling in the signaling problem, both high- and low-skilled workers could select low education. Another important characteristic of this distribution is that it can be roughly reproduced by a Poisson approximation. To show this,

Figure 6: Distribution of workers in market type 1 (managerial and professional occupations)
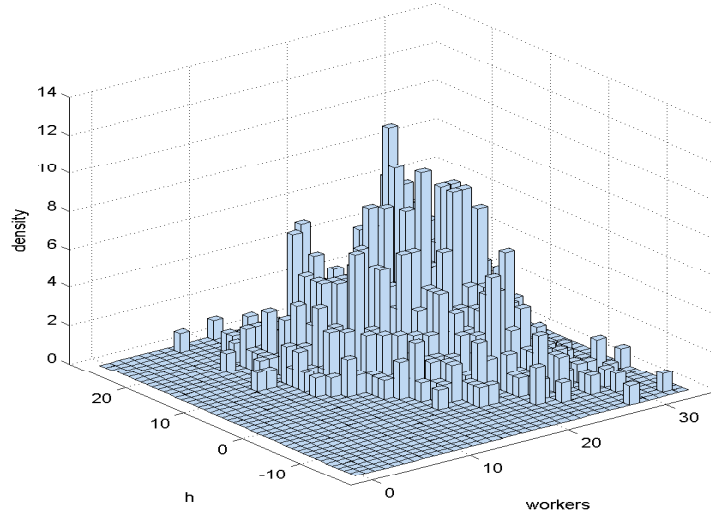

the second panel in the figure displays the bivariate distribution of two variables distributed Poisson with parameter equal to the simulated mean of workers and jobs. The similarities between the simulated and the limiting distribution are apparent.

6 Results from the quantitative model

In this section, we describe the main exercise in the paper: quantifying the impact of an improvement in the signaling quality of education on the economy.

Table 12 summarizes the model-generated data. In the first column we describe the case where $\sigma_\varepsilon = 0.01$, so that heterogeneity in costs is irrelevant. To reach perfect separation with the fixed level of education we also changed $t = -0.85$. Under such a parametrization the model provides a pure separating equilibrium in which all skilled workers decide to invest in education. Total income for skilled workers is higher than that of the unskilled. This difference is also related to the probability of unemployment: there is practically no unemployment among skilled workers while unemployment among the unskilled is about 16% -which follows from the fact that skilled workers are first in the queue in any market. Additionally, these workers have higher payoffs for reallocating between markets, because of a higher expected income compared to the cost of reallocation (for skilled workers, $\overline{W}_e/\Gamma \simeq 40$, while for the unskilled it is 33). Then, skilled workers reallocate more often. This is a key point in what follows:

Figure 7: Distribution of workers in market type 3 - Comparing simulated and
Poisson distributions

(a) Simulated model

(b) Poisson approximation
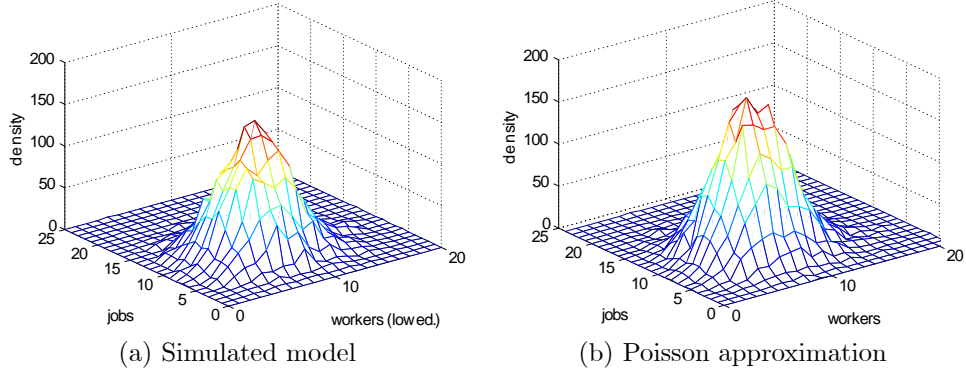
Notes: These figures compare the calibrated model distribution of workers and its Poisson
limits.

in the separating equilibrium, workers tend to allocate in those markets with
higher payoffs for their skills.

The distribution of workers in market types is shown in Figure 8. Recall that
a planner would allocate all skilled workers to market type 1, and all unskilled
workers to market type 2. The simulation results show that skilled workers
end up in market types 1 and 2, but not in type 3. Unskilled workers are still
misallocated, since they are distributed among the three market types as if at
random. As we mentioned, the gain from reallocation is not enough for them
to change occupations, mostly because the proportion of markets with poten-
tially high payoffs is relatively low (20%). Noticeably, 90% of the unemployed
unskilled are in market type 3.

We now compare the separating equilibrium to a partially pooling one in the
simulated model. Pooling arises when education is a poor signal of skills. In
our calibrated model, heterogeneity in costs is high -$\sigma_\varepsilon = 1.25$ when $t = 0$. This
case corresponds to the second column in Table 12. There, only 26% of skilled
workers and 14% of unskilled reach higher education. Thus, poor signaling
implies lower incentives to educate for skilled workers. In this calibration, this
implies lower incentives to educate overall.

While income for each educational group improves, total income does not
because of the composition effect: there is less education overall with pooling.
In fact, total expected income goes down by 15%. This follows from the fact
that in a (partial) pooling equilibrium, some of the highly skilled do not provide
any signal of their abilities and end up receiving lower wages. Additionally,

31

even though the unskilled do not invest in education, their wages rise because they cannot be separated from those skilled that have no education. Expected productivity given a low level of education is then higher.

An important outcome of our quantitative model is that misallocation of human capital is stronger in the pooling equilibrium. The first indication of this is that unemployment is higher (11% in partial pooling, as compared to 8% in separation) and that the unemployment among the skilled is now high (10%, as compared to 0%). Misallocation is also stronger in the pooling case, since many of the highly skilled end up in markets in which their skills are not as productive. For example, highly skilled workers get stuck in type 3 markets as shown in Figure 8.

Incentives to reallocate are key to explain this result. If a worker cannot signal his ability, then the payoff of reallocating to a new market is lower, because the wage after doing so would be lower and the probability of unemployment is higher. In both cases, he will be indistiguishable from an unskilled worker. This lower payoff for reallocating worsens the misallocation problem.

Overall, poor signal quality generates lower productivity as compared to that attained if signaling is perfect, due to human capital misallocation. The actual productivity drop is 24% in our calibrated model.

Table 12: Results from the simulated model

|	Separation	Pooling
Education		
% of high education	50	20
% of high skilled with educ.	100	65
Income value in flow terms ($r\overline{W}$)		
Total	10.0	9.8
Education	10.9	12.8
No education	9.2	9.06
Average productivity	17.9	13.6

Notes: Model generated data. Separation corresponds to $\sigma_{epsilon} = 0, t = -.85$, and Pooling to $\sigma_{epsilon} = 1.25, t = 0$.

Table 13: Results from the simulated economy - Unemployment

	Separation	Pooling
Unemployment		
Total	0.116	0.082
Education	0.005	0.003
No educ	0.135	0.163
High skills	0.100	0.003
Low skills	0.116	0.163

Notes: Model generated data.

Figure 8: Employed and unemployed workers in markets types 1 and 3


(a) Market type 1   (b) Market type 3
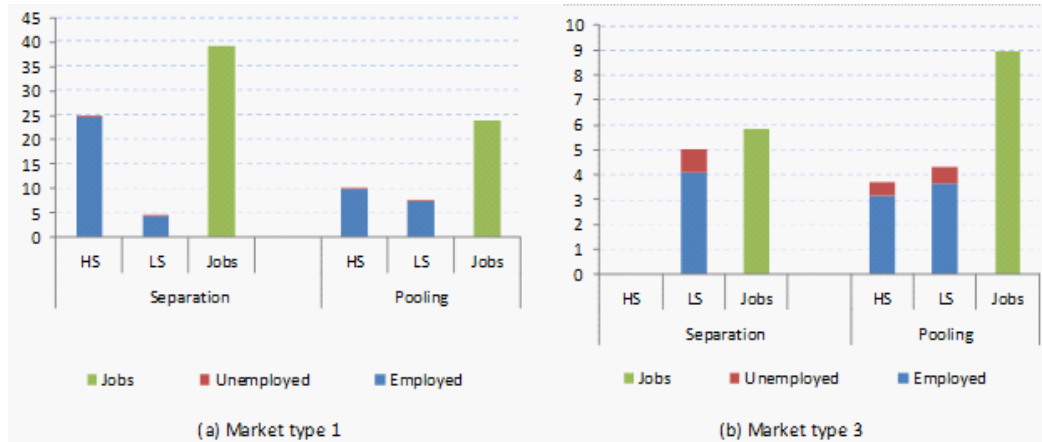
Notes: Model generated data.

Finally, to further examine effects of poor signal quality, we solved and simulated a model with a different distribution of the idiosyncratic shock to education costs, σ_ϵ. Given that our exercise implies changing both σ_ϵ and t, the parameter of the cost function for education, we computed a linear combination of both parameters, so that we change σ_ε from 1.25 to zero at the same time that we reduce t from zero to -0.85. Figure 9 reports the results of this exercise, plotting average productivity and the proportion of educated workers as a function of σ_ϵ. We find that both variables are negatively affected when we move from the separation equilibrium (low σ_ϵ) to the pooling equilibrium.

An important use of these figures is to analyze the effect of reducing, rather than eliminating, signaling problems. For example, if idiosyncratic costs halved the proportion of educated workers would rise 50% and productivity would increase about 10%.

Figure 9: Effect of poor quality of the signal
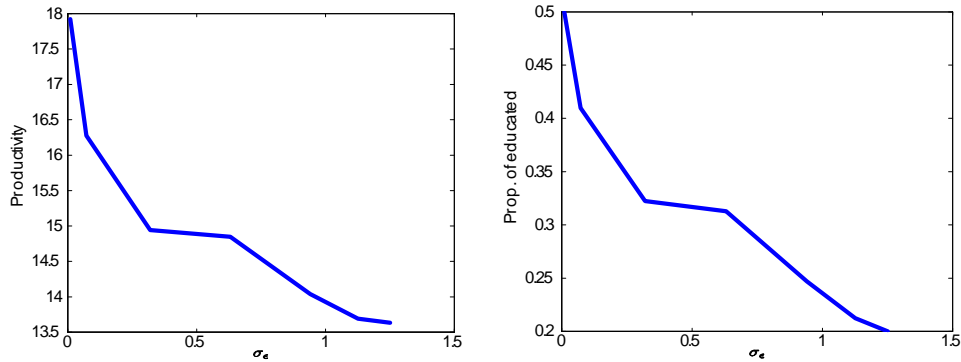
Notes: Average productivity and proportion of workers with high education in the model after changing σ_ϵ and t.

It is important to notice that in our model, as in any pure signaling model of education, education does not impact productivity directly. The link between them goes exclusively through allocation: better signaling implies easier separation between high and low skilled workers and, thus, better allocation. Thus, the positive link between education level and productivity, implied in Figure 9 and made explicit in Figure 10, should be interpreted just as the effect of signal quality and its implications on worker decisions.

Significantly, the between countries relationship between education and productivity has usually been considered within the human capital model. This paper shows that the signaling model could also be relevant to explain this relationship.

Figure 10: Education and productivity
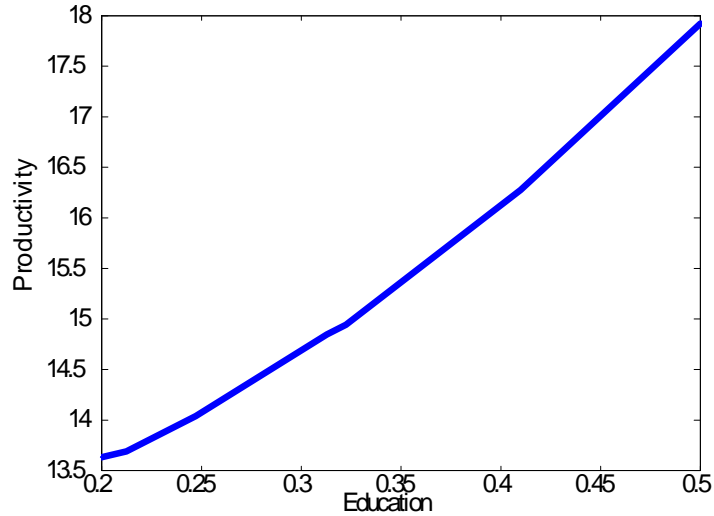
7   Discussion

In this section we discuss the relevance of the model and the extent to which it may explain the problem of low education levels, low quality, worker misallocation and low productivity in Latin America. In doing so, we describe a few aspects of the model and its quantitative implementation, qualifying some of our results. We begin with more specific issues to then focus on discussing the general interpretation of the model and its policy implications.

For the quantitative analysis, the importance of the signaling model should be identified. There are two main outcomes that help in this identification: (i) the distribution of skills within education levels and (ii) wage regressions: how much is still explained by education even after controlling for cognitive and noncognitive skills. Arguably, firms can be able to observe other signals (for example, the quality or reputation of the university a worker has attended, grades, and can even spend resources in interviewing workers). Thus, some of the skills dispersion within an educational level can, eventually, be observed by firms. If such heterogeneity were to be easily identified by employers, then wages would be less related to education and much more related to skills. In that sense, the proportion of wages explained by education after controlling for skills is an important way of identifying signaling quality.

Of course, our identification strategy rests on the observability of skills. In the data, these observations are derived from noisy tests. While we try to reduce this noise with the standard techniques of factor analysis, it would still tend to

35

increase the signaling problem in our model. While we cannot correct for that noise, this fact also highlights the importance of signaling. The difficulty for the researcher to clearly measure cognitive and noncognitive skills with specifically designed instruments also helps emphasize the problem of any employer that wants to identify the same abilities.

Our model simplifies the analysis by assuming spot markets. In such a case, workers cannot attach to any particular firm, so that the only information the firm ever has for each worker is his education level. Needless to say, if there were contracts and stable relationships between firms and workers, and if individual productivity were to be observed on the job, the information problem would be reduced with tenure and experience. This point implies that our model should be interpreted as a framework more relevant for young workers, without any other signal but education.

In our setup, the signaling problem is introduced through heterogeneity in costs of education given skills. This modelling choice provides tractability but is still ad-hoc. Alternative formulations, such as different productivity values given abilities once education is chosen, are possible, and closely related, but less easily related to the data. We assume that productivity in one market only depends on skills, which are observable characteristics. If productivity were to depend on a shock (beyond skills), it would be difficult to identify in the data.

From the quantitative exercise we argue that the signaling problem could be important, with strong consequences in unemployment and misallocation. Education, though, does not consist only on signaling and selection. The role of education is also to accumulate human capital. Our work concentrates on the effect of the signaling problem. Arguably, some of its effects could be larger if human capital accumulation were added to the process: if a signaling problem reduces the incentives to educate, it would not only generate misallocation but also lower productivity levels for the average worker.

Our analysis argues that education should provide clear signals of abilities for workers to be able to gain incentives to allocate in the correct career path. This is important, in particular, for high school education, which provides two signals: to labor market and to the unmodelled college decision and outcome.

The quantitative effects might depend on the calibration. Additionally, the particular modeling of the labor market can also reduce the effects of the analysis. To be sure, the mismatch model implies some misallocation even with perfect signaling. Moreover, firms' profits do not necessarily rise with separation as compared to pooling, which implies that the number of jobs could be lower in separation -again, as compared to pooling. To provide an intuition for this last observation, consider first complete pooling with a given, exogenous worker allocation (which may follow from Γ taking a very large value). In such a case, productivity would be the average of both types of workers, and positive profits equal to productivity minus home production (i.e. unemployment) would be probable. Consider now an exogenous allocation, but with complete separation. In that case, while productivity of educated workers would be higher, the probability of unemployment of this group would tend to be lower just because it is now a smaller group, and thus, profits could be lower. Lower profits would

imply lower job creation. For these reasons, the mismatch model could then reduce the positive effects of improving signal quality.

8   Conclusions

Through a model that combines labor markets and signaling, we explore the implications of worsening the quality of education as a signal of skills on investment in education and worker allocation.

Our qualitative analysis of the signaling model provides important insights. In particular, and given higher wages for higher skills, reducing the quality of education implies that a lower proportion of skilled workers would choose to educate. The intuition is that with a poor quality signal, many unskilled would be willing to invest in education; for that reason, the educational level required to signal high skills tends to rise endogenously and only a proportion of skilled workers, those with lower educational costs, would be willing to invest so much in the signal.

In that case, when most skilled workers choose not to educate, then the possibilities of misallocation are high. Given that workers cannot be separated according to their skills, firms cannot choose to employ first highly skilled workers (and pay them accordingly, so as to attract them), which implies that some skilled workers could be unemployed while some unskilled could be employed. Consequently, if skilled workers could end up unemployed in a market even if they are much more productive for that job, incentives to reallocate into markets in which they are more productive fall.

These effects are present in our calibrated model. In particular, we focus on Peruvian data moments, with available information on cognitive and noncognitive skills and education for a sample of the working age population. Trough this calibration we find that misallocation is strong and quantitatively substantial: poor signaling quality of education generates a productivity that is about 20% lower in the model economy, and causes unemployment to rise by two percentage points.

It is important to emphasize that our quantitative results should be analyzed as a plausible example, rather than as a precise measure, both due to the specificity of the labor market model we have adopted, and because some of these results rest on the model's calibration.

In any case, low quality in education, both as a builder and as signal of skills, seems to be relevant to understand the relatively lower development of Latin American countries. Both from our qualitative and quantitative results, we find that lower quality of education could lead to low skilled workers in higher education levels, to lower investment in education by skilled workers and to a significant misallocation of human resources.

References

[1] Allen, J. and R. van der Velden, "Educational Mismatches vs. Skill Mismatches: Effects on Wages, Job Satisfaction and On-the-Job Search," *Oxford Economic Papers*, 53, 434-452, 2001.

[2] Araujo, A., D. Gottlieb and H. Moreira, "A Model of Mixed Signals with Applications to Countersignalling," *RAND Journal of Economics*, 38, 1020-1043, 2007.

[3] Kelly Bedard, "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy*, University of Chicago Press, vol. 109(4), pages 749-775, August, 2001.

[4] Buera, F. J., J. P. Kaboski, and Y. Shin (2011). Finance and Development: A Tale of Two Sectors. *American Economic Review 101*(5), 1964–2002.

[5] Carlsson, H. and S. Dasgupta, "Noise-Proof Equilibria in Two-Action Signaling Games," *Journal of Economic Theory*, 77, 432-460, 1997.

[6] Chevalier, A., "Measuring Over-education," *Economica,* 70, 509–531, 2003.

[7] Cohn, E., Kiker, B. and Mendes De Oliveira, M. "Further evidence on the screening hypothesis", *Economics Letters*, 25, pp. 289-294, 1987.

[8] D'Erasmo, P. N. and H. J. Moscoso Boedo (2012). Financial structure, informality and development. *Journal of Monetary Economics 59*(3), 286–302.

[9] de Haan, T. , T. Offerman and R. Sloof, "Noisy Signaling: Theory and Experiment," Games and Economic Behavior, 73, 402-428, 2011.

[10] Desjardins, R. and K. Rubenson, "An Analysis of Skill Mismatch Using Direct Measures of Skills," OECD Education Working Paper 63, OECD, 2011.

[11] Erosa, A. and L. Allub (2012). Financial Frictions, Occupational Choice, and Economic Inequality. 2012 Meeting Papers 702, Society for Economic Dynamics.

[12] Feltovich, N., R. Harbaugh and T. To (2002), "Too Cool for School? Signaling and Countersignaling," *RAND Journal of Economics* (33), 630-649.

[13] Groot, Wim and Oosterbeek, Hessel, "Earnings Effects of Different Components of Schooling: Human Capital versus Screening," *The Review of Economics and Statistics*, MIT Press, vol. 76(2), pages 317-21, May, 1994.

[14] Heckman, J. and T. Kautz, "Hard Evidence on Soft Skills," *Labour Economics*, 19, 451-464, 2012.

[15] Heckman, J., J. Stixrud and S. Urzua (2006) "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior", *Journal of Labor Economics*, 24, 3, 411-482, 2006.

[16] Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics 6*(1), 735–770.

[17] Jaeger, David A and Page, Marianne E, "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *The Review of Economics and Statistics*, MIT Press, vol. 78(4), pages 733-40, November, 1996.

[18] Jovanovic, B. (1979), "Job Matching and the Theory of Turnover,", *Journal of Political Economy* (87), 972-990.

[19] Katz, E. and Zimmerman, A. "On education, screening and human capital", *Economics Letters*, 6, pp. 81-88. 1980.

[20] Kautz, T., J. Heckman, R. Diris, B. ter Weel and L. Borghans, "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success," NBER working paper, 2014.

[21] Kevin Lang and David Kropp, "Human Capital Versus Sorting: The Effects of Compulsory Attendance Laws," *The Quarterly Journal of Economics*, Oxford University Press, vol. 101(3), pages 609-624, 1986.

[22] Layard, Richard and Psacharopoulos, George, "The Screening Hypothesis and the Returns to Education," *Journal of Political Economy*, University of Chicago Press, vol. 82(5), pages 985-98, Sept./Oct, 1974.

[23] MacLeod, W., E. Riehl, J. Saavedra and M. Urquiola, "The Big Sort: College reputation and labor market outcomes". NBER working paper 21230, 2015.

[24] Matthews, S. and L. Mirman, "Equilibrium Limit Pricing: the Effects of Private Information and Stochastic Demand," *Econometrica*, 53, 981-996, 1983.

[25] Pellizzari, M. and A. Fichen, "A New Measure of Skills Mis-match: Theory and Evidence from the Survey of Adult Skills (PIAAC)," OECD Social, Employment and Migration Working Paper 153, OECD, 2013.

[26] María F. Prada and Sergio S. Urzúa "One Size does not Fit All: Multiple Dimensions of Ability, College Attendance and Wages," NBER Working Paper 20752, National Bureau of Economic Research, 2014.

[27] Riley, John G., "Testing the Educational Screening Hypothesis," *Journal of Political Economy*, University of Chicago Press, vol. 87(5), pages S227-52, October, 1979.

[28] Riley, John G., "Silver Signals: Twenty-Five Years of Screening and Signaling," Journal of Economic Literature, 39, 432-478, 2001.

[29] Shimer, R., "Mismatch," *American Economic Review*, 97, 1074-1101, 2007.

[30] Spence, M. *Market Signalling: Information Transfer in Hiring and Related Processes*, Harvard University Press, 1973.

[31] Sicherman, N., "Overeducation in the Labor Market," *Journal of Labor Economics*, 9, 101–122, 1991.

[32] Tyler, John H., Richard J. Murnane and John B. Willett, "Estimating the Labor Market Signaling Value of the GED," *The Quarterly Journal of Economics*, Oxford University Press, vol. 115(2), pages 431-468, 2000.

[33] Weiss, Andrew, "Human Capital vs. Signalling Explanations of Wages," *Journal of Economic Perspectives*, American Economic Association, vol. 9(4), pages 133-154, Fall, 1995.

[34] Wolpin, K. I., "Educational screening" American Economic Review, 67, 949-958. 1977.

[35] Yamada, G., P. Lavado, and L. Velarde (2014), "Habilidades no cognitivas y brecha de género salarial en el Perú," *Estudios Públicos,* 0, 89–129, 2014.

Appendix

Data for calibration

In this appendix, we explain how we compute the moments we select for the calibration of our model.

Our data sources are ENHAO and ENHAB from Peru. In both cases we keep only individuals aged between 23 to 40. We define a unique skills indicator by running a factor model on all variables regarding skills, cognitive and noncognitive abilities, and select only one factor. We define as highly skilled those observations above the 50th percentile in this skills factor. In addition, we consider as educated those who have completed tertiary education or college. The proportion of workers in each group, by skill level and education are used as calibration targets.

We also resort to wage regression results. In particular, the last column of Table 9 show the moments we use.

So as to analyze transitions, we use ENAHO (restricting to the 2010-2011 panel). We compute transitions from unemployment to employment and transitions from employment to unemployment.

Additionally, we find that 15% of workers changed occupations (broadly defined in three categories) within a year. Among those that changed jobs, about 40% changed occupations.

We convert all transition rates to quarterly transition rates and use them as moments in the calibration.