



Cognitive and socioemotional skills and wages: the role of latent abilities on the gender wage gap in Peru

Pablo Lavado ¹ · Luciana Velarde¹ · Gustavo Yamada¹

Received: 17 September 2019 / Accepted: 15 February 2021 / Published online: 2 April 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The literature provides evidence on the positive connection between cognitive test scores and higher wages. Fewer and newer studies have explored the correlation between non-cognitive test scores and wages. However, these studies only focus on developed countries. The main objective of this study is to identify latent abilities and explore their role in the gender wage gap in a developing country: Peru. The main identification strategy relies on exploiting panel data information on test scores and arguing that time dependence across measures is due to latent abilities. We exploit two databases: the Young Lives Study and the Peruvian Skills and Labor Market Survey. The results show that when accounting for differences in actual latent abilities, socioemotional abilities account for important inter-gender differences in the endowment and returns of abilities. Moreover, inter-gender differences in latent abilities play an important role in not only wage profiles but in schooling, employment, and occupational decisions.

Keywords Cognitive skills · Socioemotional skills · Gender wage gap

JEL codes J16 · J24

1 Introduction

Differences in cognitive skills are strongly related to differences in wages between females and males (Murnane et al. 1995; Murnane et al. 2000; Grogger and Eide 1995; Weinberger 1999; Altonji and Blank 1999). Specifically, males have higher cognitive test scores than females on average (Strand 2003; Bell et al. 2006; Upadhayay and Gurugain 2014) and these scores contribute to higher wages for the

✉ Pablo Lavado
p.lavadopadilla@up.edu.pe

¹ Universidad del Pacífico, Jr. Sánchez Cerro, 2050 Lima 11, Peru

former than for the latter (Hedges and Nowell 1995; Paglin and Rufolo 1990). Furthermore, differences in cognitive skills could contribute to the gender wage gap not only because of differences in means but also because of differences in returns: scoring an additional point on a cognitive test results in a larger gain in terms of wages for a male than for a female.

Recently, studies have focused on the relation between socioemotional skills and productivity. The main finding is that there is a positive connection between wages and certain socioemotional skills. Early work by Marxist economists showed that employers in low-skill labor markets value docility, dependability, and persistence more than cognitive ability or independent thought (see Bowles and Gintis (1976) and Edwards (1976)). Recent studies (see Heckman et al. (2006), Cunha and Heckman (2007), Hanushek and Woessmann 2008) also support this fact with evidence of a positive relation between results in socioemotional test scores and labor market outcomes such as wages or occupational choice.

Regarding how people form socioemotional skills, some studies propose that test scores are bad proxies of abilities due to measurement errors and endogeneity with schooling. The important features that drive wages are latent abilities. However, to the best of our knowledge, few studies exist that address the role of latent socioemotional and cognitive abilities in the gender wage gap. Moreover, they do not address this question for the developing world and for Latin American countries in particular.

Based on the latent ability model proposed by Heckman et al. (2006), we estimate the contribution of socioemotional skills to the gender wage gap. In contrast to Heckman et al. (2006), our identification strategy is based on panel data collected by the Young Lives Study (YL) for Peru. We argue that dependence through time between test scores is due to latent abilities. After estimating latent abilities, we use them in an Oaxaca–Blinder decomposition to explore the role that abilities and their returns play in explaining the gender wage gap. Moreover, we estimate a joint model of schooling, employment, occupational choice, and earnings to disentangle the effect of inter-gender differences on the ability to make each of these choices. Since the YL database lacks information on wages, we estimate latent abilities as linear combinations of characteristics common to both YL and Peruvian Skills and Labor Market Survey (ENHAB). The latter is a recent survey in Peru which gathers data on cognitive and socioemotional test scores on an individual's characteristics, educational trajectory, and wages.

The preliminary results show that there is a significant gender wage gap in Peru. In fact, in a model with measured abilities, we find significant inter-gender differences in the endowment of cognitive skills but no relevant differences in terms of socioemotional abilities (in endowment or returns). Estimating the joint model shows that differences in socioemotional abilities between men and women are important but only to the choices prior to determining wages. Cognitive skills are relevant in determining years of schooling and occupational choice and measuring socioemotional ability for wages and employment. Applying our proposed estimation procedure shows that the actual latent ability turns out to be highly and statistically significant for mean wages as well as accounting for inter-gender differences. In particular differences in the endowment of socioemotional abilities contribute (negatively) to the gender wage gap. However, we do not find any significant

inter-gender differences in the returns to cognitive abilities. Moreover, the estimation of the joint model sheds light on the fact that the observed gender wage gap is mainly attributable to differences in occupational choices. Cognitive and socioemotional abilities have different values for men and women in terms of schooling, employment, and wages; but basically men earn higher wages because their equilibrium assignation is to occupations with higher rewards for cognitive skills that they are most endowed with.

The paper is organized as follows: the following section presents a review of the literature. Section 3 presents our empirical baseline model of wages in terms of abilities as well as the Oaxaca–Blinder decomposition. In Section 4, we describe the data and sample. Section 5 presents our econometric implementation of estimating latent abilities. Section 6 presents the results. And, Section 7 concludes.

2 Literature review

For decades, researchers have focused on studying the relation between abilities and labor market outcomes. These studies are mostly related to cognitive test scores. Murnane et al. (1995) assess the role of the mathematics skills of graduating high school seniors in their wages at age 24 and find a positive and increasing effect of cognitive skills on wages (especially in years closer to graduation). In a more recent study, Cunha et al. (2006) state that cognitive ability affects the likelihood of acquiring higher levels of education and advanced training as well as the economic returns to these activities. Other studies have focused on explaining the black-white wage gap, and they confirm the importance of the cognitive factor: large observed cognitive gaps between black and white workers in their late twenties are an important determinant of the wage gap (Neal and Johnson 1996) and unobserved cognitive ability is the most important variable in explaining the racial differences in wages (Urzúa 2008).

Somewhat more recently, studies have directed their attention towards socio-emotional abilities and their relation to labor market outcomes. Early work by Bowles and Gintis (1976), Edwards (1976), and Klein et al. (1991) show that socioemotional skills such as dependence and persistence are highly valued by employers. Other studies such as that of Heckman et al. (2006) also support this fact with evidence of a positive relation between results in socioemotional test scores and outcomes in the labor market. More recently, a type of non-cognitive abilities that has received attention are the social ones. Specifically, Deming (2017) finds that a pre-market measure of self-reported sociability, the number of clubs and sports in which the respondent participated in high school, is a significantly more important predictor of labor market success for this last cohort than for the first one: the association between social skills and the probability of full-time work increased more than fourfold and the return to social skills among full-time workers grew significantly. Additionally, as in Weinberger (2014), Deming (2017) finds that cognitive and social skills are complements but does not find complementarity between cognitive skills and the widely used measures of non-cognitive skills.

The research that explores gender inequality at first focused on studying the role of cognitive skills in explaining gender differences on market labor outcomes. Paglin

and Rufolo (1990) report that most of the gender gap in average starting salaries for college graduates is between, rather than within, detailed college majors and that differences in starting salaries across majors have a very positive relation with average math scores. Moreover, the gender difference in math scores accounts for around 20% of the gender gap in the starting salaries of college graduates. Likewise, Landes (1977), Ragan and Smith (1981), and Mincer and Ofek (1982) estimate that the accumulation of human capital, the occupational selection, and turnover account for a large portion of the gender wage gap.

However, because of a sizeable component in the gender wage gap continues to go unexplained, researchers started reaching out beyond the confines of traditional economic models for explanations. One of these is the hypotheses that psychological attributes or non-cognitive skills can explain gender differences in the labor market. The evidence indicates that women are more risk averse than men, that women may systematically underperform relative to men in competitive environments, and that women have a higher level of altruism and stronger preferences for redistribution (Bertrand 2011). Despite this, Blau and Kahn (2017), with respect to the possible psychological gender differences, make some clarifications: first, even if men and women do differ on average, it is not possible at this point to know the role of nature versus nurture; and second, gender differences in non-cognitive skills do not necessarily all favor men.

Some authors detect considerable effects from socioemotional abilities on the gender wage gap. For instance, Grove et al. (2011) find that personality traits and preferences regarding family, career, and jobs account for a quarter of the “explained” gender wage gap. Similarly, Fortin (2008) finds that socioemotional factors account for a small but nontrivial part—about 2 log points— of the gender wage gap of workers in their early thirties. This magnitude is larger than that of educational attainment and cognitive skills (math scores), which account for about 1.2 log points, and almost as important as that of labor market experience and job tenure, which account for 2.4 log points.

Nonetheless, other research finds modest effects of socioemotional abilities. Manning and Swaffield (2008), in respect to the gender gap in early career wage growth in the UK, conclude that the whole set of psychological factors they use— comprised of risk aversion, competitiveness, self-esteem, other-regarding behavior, and career orientation—can explain at most 4.5 log points of the about 25 log point gap in earnings that has built up between men and women 10 years after entering the labor market, and human capital factors account for about 11 of these 25 log points. Hence, as Bertrand (2011) highlights, the research on the effect of gender psychological differences is clearly just in its infancy and far from conclusive and has many contradictory findings.

On another note, quite remarkable is the fact that most of the research regarding the role of cognitive and socioemotional test scores in the labor market outcomes has only focused on developed countries. To the best of our knowledge only a few address this issue in developing countries, in particular in Latin American countries.

Díaz et al. (2012) use the ENHAB on a sample of the working age population and apply an instrumental variable approach to address issues regarding the endogeneity from schooling. They find that schooling and both cognitive and socioemotional skills are valued in the Peruvian labor market: one standard deviation increase in the

years of schooling generates an increase of 15% in earnings, while a change in cognitive skills and socioemotional skills of a similar magnitude generate a 9 and 5–8% increase in earnings, respectively. However, their strategy considers that skills only affect wages directly and through schooling. Unlike us, they do not consider that skills also affect the decision to work and the kind of occupation chosen. Moreover, Urzúa et al. (2009) go a step further by analyzing gender discrimination in the labor market of Chile with a rich data set. They follow the studies that estimate labor market models with multiple sources of unobserved heterogeneity through cognitive and socioemotional abilities. Nonetheless, due to data limitations, they consider only one underlying source of unobserved heterogeneity as a combination of cognitive and socioemotional abilities. Their results show the existence of gender gaps in the labor market variables such as experience, employment, hours worked, and hourly wages that cannot be explained by observable or unobservable characteristics, or the underlying selection mechanisms that generate endogeneity. As far as we know, this is the only study that addresses the role of cognitive and socioemotional (latent) abilities (accounting for the endogeneity from schooling) in the gender wage gap in a developing country. So, our attempt to uncover the importance of latent abilities to the gender wage gap will be one of the first on developing countries.

Other ways to identify the effect of abilities on labor outcomes are by using fixed effects estimates or relying on instrumental variables. A benefit of using twin fixed effect estimates is that they control for genetic factors and background perfectly. Thus, there is no need to rely on test scores to control by the ability. However, we do not have a dataset that lets us exploit this type of fixed effects. We do have information on test scores.

The problem with test scores is that they do not reflect the real (latent) abilities because they have a measurement error (Heckman et al. 2006). Therefore, using them in wage and schooling regressions is troublesome. Conditioning on schooling, both cognitive and socioemotional tests predict wages. However, schooling is a choice variable and thus its endogeneity must be addressed. Omitting schooling from the wage equation increases the correlation of both abilities with wages. Estimates comprise both the direct (on productivity) and indirect (on schooling) effects of abilities on wages. Nonetheless, there is an important difference between cognitive (and socioemotional) tests and achievement tests. Although IQ is well set by age 8, the research has demonstrated that achievement tests are quite malleable and increasingly so with schooling. This malleability creates a reverse causality problem.

To overcome this, Hansen et al. (2004) develop two methods for estimating the effect of schooling on achievement test scores that control for the endogeneity of schooling by postulating that both schooling and test scores are generated by a common unobserved latent ability. Their analysis shows that schooling has small equalizing effects on measured test scores, especially for those with low ability and low levels of schooling. In the same line, Behrman et al. (2008) find that the rates of return are much higher for investing in primary school than for investing in middle schools and at the primary school level, the returns are higher for expanding low-quality schools than for increasing the quality of existing schools.

More recently, Behrman et al. (2011) estimate the effect of cognitive skill and physical health on wages. Their results, when accounting for this endogeneity, show that only the cognitive skill has an impact on wages; health does not. This result opposes earlier evidence that showed positive associations between health human

capital and wages. In this respect, we are aware that wages and other labor outcomes are not only driven by abilities, but the physical characteristics and health of an individual also affect productivity. Unfortunately, even though we acknowledge the importance of health to productivity, we cannot contribute to this discussion due to data limitations.

To summarize, the main contribution of our analysis is exploring the role that latent cognitive and socioemotional abilities and their returns in terms of wages play in explaining the gender wage gap in a developing country. We achieve this by using a data combination method and by modeling the endogeneity present in the job decision process.

3 Model

The model is based on Heckman et al. (2006), Cunha et al. (2010), and Cunha and Heckman (2008). Latent cognitive and socioemotional abilities are two underlying factors. Conditioning on observables, these factors explain all the dependence across choices and outcomes. Individuals make decisions regarding schooling, working, and occupation. If the individual works, he or she earns a wage.

3.1 The model for wages

As in Heckman et al. (2006), we let f^C and f^S denote the latent cognitive and socioemotional abilities, respectively, and assume they are independent. The logarithm of wages are given by:

$$\ln W = \beta_Y X_Y + \alpha_Y^C f^C + \alpha_Y^S f^S + e_Y \quad (1)$$

where X_Y is a vector of observed controls, β_Y is the vector of returns, α_Y^C and α_Y^S are the latent cognitive and socioemotional abilities, respectively, and e_Y represents the error term. In order for Eq. (1) to be identified, we assume that the error term is independent of all other factors. However, unobserved factors that affect wages and productivity, such as health, could be correlated with cognitive and non-cognitive skills. In this context, we are cautious in interpreting α_Y^C and α_Y^S as causal effects. Instead of estimating the total effect, our analysis focuses on estimating the contribution of skills in the variation among wages.

The identification strategy is similar to that in Heckman et al. (2006). We restrict the latent cognitive ability to only affecting cognitive measures and the latent socioemotional ability to only affecting socioemotional measures. The model of the cognitive measure is:

$$C = \beta_C X_C + \alpha_C f^C + e_C \quad (2)$$

Likewise, the model of the socioemotional measure is:

$$S = \beta_S X_S + \alpha_S f^S + e_S \quad (3)$$

Our assumptions indicate that conditional on the X variables, the dependence across time of measurements comes from f^S and f^C .

3.2 The model for schooling

Each individual chooses the level of schooling that maximizes his or her lifetime expected benefit. Following a linear-in-the-parameters specification, and letting I_{S_c} represent the net benefit associated with schooling level S_c , we have:

$$I_{S_c} = \beta_{S_c} X_{S_c} + \alpha_{S_c}^C f^C + \alpha_{S_c}^S f^S + e_{S_c} \quad (4)$$

where S_c is the schooling level chosen by the individual among $\overline{S_c}$ possibilities; X_{S_c} is a vector of observed variables that affect schooling; β_{S_c} is its associated vector of parameters; $\alpha_{S_c}^C$ and $\alpha_{S_c}^S$ are the factor loadings associated with cognitive and socioemotional latent abilities, respectively; and e_{S_c} represents an idiosyncratic component assumed to be independent of f^S , f^C , and X_{S_c} . The error terms for each schooling level are mutually independent.

The observed schooling level corresponds to:

$$D_{S_c} = \operatorname{argmax}_{S_c \in 1, \dots, \overline{S_c}} [I_{S_c}] \quad (5)$$

We consider two educational levels: (i) complete tertiary education, and (ii) up to complete secondary education. We consider that schooling's relevant decision is whether to complete tertiary education or not, so that this decision is consistent with the occupational choice of a high-skilled or low-skilled job (Section 3.4). We use an indicator variable $D_{S_c} = 1(I_{S_c} > 0)$ to indicate the choice of attaining complete higher education.

3.3 The model for employment

Let I_E denote the net benefit associated with working and assuming a linear-in-the-parameters specification:

$$I_E = \beta_E X_E + \alpha_E^C f^C + \alpha_E^S f^S + e_E \quad (6)$$

where β_E , X_E , α_E^C , α_E^S , and e_E are defined as in the schooling model. Then we observe whether the individual is employed which corresponds to a binary variable $D_E = 1(I_E > 0)$ that equals one for employment and zero otherwise. The error term is orthogonal to the control variables.

3.4 The model for occupational choice

Let I_0 denote the utility that is associated with choosing a white collar occupation (where the alternative is a blue collar occupation). We are not using the conventional definition of white and blue collar work (type of occupation). White collar is working in a job that requires higher skills. We are defining this high-skilled job as one that requires a tertiary education. This definition makes sense for Peru where skilled labor is scarce: only 15% of individuals graduate with a tertiary education (Ministry of Education of Peru). We assume the following linear model for I_0 :

$$I_0 = \beta_0 X_0 + \alpha_0^C f^C + \alpha_0^S f^S + e_0 \quad (7)$$

where β_0 , X_0 , α_0^C , α_0^S , and e_0 are defined as in the schooling and employment models. $D_0 = 1(I_0 > 0)$ is an indicator of the choice of white collar occupational status. The error term is orthogonal to the control variables.

Both cognitive and non-cognitive factors are known by each individual but not for the econometrician. Controlling for this dependence is equivalent to controlling for the endogeneity in the model. The wage equations usually are functions of measured abilities or test scores. However, these test scores have a measurement error and are functions of schooling and latent abilities. For that purpose, using measured abilities does not reflect the parameters associated with the effect of abilities on choices and labor market outcomes.

In order to deal with this endogeneity problem, Heckman et al. (2006) estimate the distributions of latent abilities that rely on having at least three measurements. In addition to that, our identification strategy also relies on having panel data information on measurements; specifically, having information on the same measure in two different moments in time. Finally, even though we are assuming a linear-in-the-parameters specification, the model can be interpreted as an approximation of a more flexible behavioral model as in Heckman et al. (2006).

4 Data and sample

The identification strategy relies on having panel data information on the measured abilities, schooling, labor force participation, occupational choices, and wages. Unfortunately, this database is not available in Peru. We propose an empirical method which exploits two datasets.

The first one is the YL database for Peru. The YL contains longitudinal information on two cohorts of children (younger cohort and older cohort) for each of four countries: Ethiopia, India (Andhra Pradesh), Peru, and Vietnam. In Peru, data were collected from 20 sites in 14 regions that represent 95% of the Peruvian child population (excluding the 5% with higher incomes). Children and their caregivers were interviewed three times: in 2002 (baseline survey), when they were 8 years old; in 2006–2007, when they were 12 years old; and again in 2009–2010, when they were around 15 years old. The survey contains information on aspects related to child development, cognitive test scores, psychosocial traits (attitudes and aspirations), and anthropometric measures as well as a rich set of other individual and household characteristics. In particular, household characteristics such as household socio-status, wealth indices, log household consumption, and caregivers' measured ability are also shown as well as other individual characteristics.

In order to analyze the distribution of skills among Peruvian children, we focused on the older cohort which, for Peru, comprised around 700 children that were 8 years old by the beginning of the study (born in 1994–5). We worked with the subsample of children with available information on items related to cognitive and socio-emotional abilities as well as individual characteristics for rounds 2 and 3.¹ Finally, we worked with the subsample of children living in urban areas. The final sample comprised 349 individuals. The children in our sample and those urban children whose full information we were not able to find in rounds 2 and 3 came from households with similar wealth and consumption per capita and were, on average, of

¹ Information regarding socioemotional abilities was not collected during the first round for the older cohort.

Table 1 Descriptive statistics (young lives)

	Round 2				Round 3			
	Mean	SD	Female	Male	Mean	SD	Female	Male
Cognitive ability (PPVT raw)	76.92	13.84	75.964	77.777***	101.083	14.599	99.788	102.245***
Self efficacy index	0.101	0.939	0.216	-0.003***	0.101	0.969	0.295	-0.073***
Self esteem index	0.139	0.897	0.088	0.185***	0.147	0.940	0.208	0.093***
Caregiver's self efficacy	0.144	0.997	-0.012	0.284***	0.145	0.925	0.136	0.154
Caregiver's self esteem	0.046	1.003	0.058	0.036	0.044	1.028	0.130	-0.033***
Height for age	-1.302	1.049	-1.337	-1.271*	-1.306	0.863	-1.502	-1.129***
Body mass index	0.340	0.966	0.297	0.378***	0.272	0.972	0.432	0.128***
Age in months	148.867	5.416	148.488	149.206**	179.117	4.505	178.783	179.417***
Schooling	6.143	0.895	6.170	6.120	9.006	1.101	9.067	8.951
Missed school	0.054	0.227	0.055	0.054	0.069	0.253	0.067	0.071
Wealth index	0.602	0.190	0.596	0.608**	0.661	0.150	0.665	0.658
Log household consumption	5.191	0.642	5.179	5.201	5.333	0.635	5.338	5.328
Mother tongue (Spanish)	0.788	0.409	0.794	0.783				
Father's educational Level	10.481	3.092	10.176	10.755***				
Mother's educational level	9.602	3.499	9.648	9.560				
N	349		165	184	349		165	184

*Indicates 10% significance level; **Indicates 5% significance level; ***Indicates 1% statistical significance level

similar age, weight, and height. We find that mothers of children in our sample had attained more years of schooling at the time of round 1 in comparison to the mothers in the missing subsample. While the balance in general indicates that the missing information was random, we cautiously interpreted our results as valid for children with higher cognitive and non-cognitive abilities, as their parents' educational levels were important determinants of ability.²

The YL subsample were evenly distributed among boys and girls (165 and 184, respectively) with an average age of 149 months and a mean of 6 years of schooling in round 2. Table 1 presents the descriptive statistics on the main variables of interest from both rounds as well as information on the child's mother tongue and the

² For self-esteem, the statements explored in the YL survey focused on positive and negative dimensions of pride and shame based on the Rosenberg Self-Esteem Scale that focused on dimensions of children's living circumstances. For self-efficacy we focused on five items: "If we try hard we can improve my situation in life"; "Other people in my family make all the decisions about how we spend my time"; "I like to make plans for my future studies and work"; and "I (don't) have choice about the work we do". The degree of agreement was measured on a 4-point Likert scale that ranged from strong agreement to strong disagreement. We constructed two indices (one for each trait) as the average score of these items and used the standardized indices for our estimations.

parents' educational levels from round 1 (what we call "permanent characteristics"). Some facts worth highlighting are that in both rounds, while girls scored below average in items related to cognitive ability, boys did so in self-efficacy items (results were mixed for self-esteem between rounds). Most household characteristics and family backgrounds were similar between genders. Caregivers' measured socio-emotional abilities differed between children's genders; boys' caregivers showed higher levels of self-efficacy by the time of round 2 and lower levels of self-esteem by the time of round 3. Important differences appeared between both rounds of the survey; a fact that was helpful for our identification strategy. The measures used to represent socioemotional abilities were built based on respondents' degree of agreement or disagreement with a number of statements related to self-esteem. For this measure, the statements explored in the YL survey focused on positive and negative dimensions of pride and shame based on the Rosenberg Self-Esteem Scale. The degree of agreement was measured on a 4-point Likert scale that ranged from strong agreement to strong disagreement. We constructed the self-esteem index as the average score of these items and used the standardized index for our estimations.

The second database corresponds to a novel household survey collected by the World Bank in 2010 that not only contained information on wages and individual characteristics, but also on measured cognitive and socioemotional abilities for a sample of currently employed working age individuals. The ENHAB is a nationally representative household survey that comprises information on urban areas of 11,235 randomly selected individuals aged 14–65 from 2600 cities. The data contain information on household living conditions, demographic information, academic achievement, employment and earnings, and novel information on (i) cognitive and socioemotional test scores, (ii) schooling trajectories, (iii) early labor market participation, and (iv) family characteristics. The measured abilities were assessed by means of cognitive tests that evaluated numerical and problem-solving skills, working memory, verbal fluency and receptive language, and socioemotional abilities according to GRIT scales (Duckworth et al. 2007) and the Big Five personality factors (Goldberg 1990). For this analysis we focused on seven of these measures, the standardized values of each of the big-five factors (emotional stability, extraversion, agreeableness/kindness, agreeableness/cooperation, conscientiousness strong, and openness) and a compound of the two measures of Grit as well as a compound of cognitive measured abilities. Information regarding individual characteristics included personal educational background, family characteristics, and socioeconomic status (parental education and occupations, family size, information on access and school characteristics, when parents attended basic and secondary education, perceived socio-economic status, etc.).

We worked with three subsamples: (i) individuals with available information on measured abilities (test scores), $N = 2415$; (ii) individuals with positive earnings, $N = 4063$; and (iii) individuals with available information on relevant individual characteristics, $N = 7499$. In general terms, individuals in the data set were evenly distributed among men and women and had a mean age of 33 years, monthly earnings of around 1000 soles in constant Peruvian currency for year 2010 (around 350 USD), worked an average of 51 h a week, and had on average a complete secondary education. Table 2 shows some other relevant descriptive statistics for the three subsamples and the difference in each between men and women. Some facts

Table 2 Descriptive statistics (ENHAB)

	<i>N</i>	Mean	SD	Female	Male
Cognitive ability	2417	42.008	14.921	39.805	44.934***
Consistency	2417	-0.005	0.996	0.038	-0.072***
Persistence	2417	0.000	1.001	-0.049	0.057**
Grit	2417	-0.003	0.996	-0.009	-0.008
Extraversion	2417	0.002	0.999	-0.049	0.084**
Kindness	2417	-0.003	1.007	0.037	-0.055*
Cooperation	2417	0.005	0.991	0.046	-0.061**
Conscientiousness	2417	-0.004	1.001	0.060	-0.095**
Emotional stability	2417	0.012	0.996	-0.082	0.139***
Openness	2417	-0.004	0.999	-0.054	0.086*
Log hourly wage	4083	1.333	0.851	1.187	1.423
Monthly earnings	4083	1083.222	2334.846	831.617	1238.954***
Hourly earnings	4083	5.988	13.510	5.263	6.436***
Weekly work hours	4083	51.057	18.489	46.261	54.025***
Experience	4083	25.421	13.572	25.282	25.507
Age	7478	33.415	15.176	33.137	33.711**
Schooling	7478	10.705	3.380	10.524	10.898***
Mother tongue	7478	0.910	0.286	0.913	0.907
Father's education	7478	7.120	5.379	7.087	7.154
Mother's education	7478	5.731	5.288	5.658	5.808**

*Indicates 10% significance; **5% significance; ***1% significance level of the mean difference between males and females

worth highlighting are that men earn higher earnings (monthly and hourly), work longer hours, and have higher levels of measured cognitive abilities than women. These differences are statistically significant at 99%. However, the results are mixed regarding socioemotional skills. While women appear to be more consistent, kind, cooperative and conscientious, men appear to be more persistent, extraverted, emotionally stable, and open.

5 Econometric implementation

The main objective of this paper is to identify the contribution of abilities to the gender wage gap. The main equation is a function of schooling and ability:

$$\ln W_i = \alpha + \gamma S_{ci} + \gamma_{fc} f_i^c + \gamma_{fs} f_i^s + \mu_i \quad (8)$$

where $\ln W_i$ are log earnings, S_{ci} represents years of schooling, and f_i is the latent ability: Cognitive (C) or Socioemotional (S). The main problem of this equation is that f_i is unobserved by the econometrician. Thus, if schooling is correlated with ability, and ability is omitted, the estimation of γ is inconsistent. In particular, if ability is positively correlated with schooling, γ will be overestimated. The empirical

literature has dealt with this issue by including tests scores as proxies for these abilities:

$$\ln W_i = \alpha + \gamma S_{ci} + \gamma_C C_i + \gamma_S S_i + v_i \quad (9)$$

where C_i and S_i are standardized test scores for measured cognitive and socioemotional abilities. However, using test scores does not solve the problem satisfactorily. Test scores are likely to be determined not only by schooling but also by the latent abilities of the individual. Thus, the coefficient corresponding to test scores would be partially capturing the indirect effect of schooling on earnings through the measured skills; thus, the true effect of schooling on earnings cannot be obtained. Moreover, since f_i is still omitted; γ , γ_C , and γ_N are overestimated.

We propose an econometric procedure to estimate latent abilities, f_i . For that purpose, we exploit panel data information on the measured abilities from the YL database and information on wages and measured abilities from the ENHAB. Specifically, the econometric implementation is divided into four stages.

First, we use time variation (from rounds 2 and 3) in the measured cognitive and socioemotional test scores and years of schooling among children in the YL sample to recover the (unobserved) fixed effects. In particular, we try to explain the variation in the two measures of ability (MA_{it}): one socioemotional ability (self-esteem) and one cognitive ability (Peabody Picture Vocabulary Test scores). The identification procedure requires controlling for characteristics (X_{it}) that may have varied between the ages of 12 and 15 and that may explain the variation in the measured abilities during that period. In this way we can explain the changes in the measured ability and partial out any unobserved fixed effect, which we interpret as the latent ability. This latent ability collects all the information about the ability formed up to age 12.

$$\Delta MA_{it} = \gamma_0 + \gamma_X \Delta X_{it} + \Delta \mu_{it} \text{ for } i = 1, 2 \quad (10)$$

Second, we estimate the correlation of characteristics that remain unchanged in the child's life from 12 to 15 years old for these fixed effects. For this, we capture the fixed effect or unobserved component of each ability by using the first-stage estimates to predict the average value of the covariates in rounds 2 and 3 and by deviating the predicted value of the measured ability with respect to the observed value of the variable.

$$\widehat{MA}_{it} = \widehat{\gamma}_0 + \widehat{\gamma}_X X_{it} \quad (11)$$

$$\widehat{LA}_i = \frac{1}{2} \left[\left(MA_{i1} - \widehat{MA}_{i1} \right) + \left(MA_{i2} - \widehat{MA}_{i2} \right) \right] \quad (12)$$

With these estimated proxies of latent abilities, we estimate the effects of the variables that remain constant when a child is between 12 and 15 years old and which may determine the latent ability by using the YL sample. Since we are using two databases, we require that these variables are available both in the YL questionnaire as well as for the ENHAB questionnaire. This availability allows for predicting the value of the "latent ability" for the ENHAB sample, which has the information on wages. Good candidates are gender, mother tongue, and parents' educational level

(years of schooling) and are denoted as Z_i .

$$\widehat{LA}_i = \gamma_0^{LA} + \gamma_1^{LA} Z_i + \mu_i^{LA} \quad (13)$$

Third, we use the estimated parameters of the second stage to predict the fixed effects that would correspond to the ENHAB working age sample. This is possible due to the fact that the “permanent” characteristics are also available for the ENHAB sample.

$$\widehat{\widehat{LA}}_i = \widehat{\gamma}_0^{LA} + \widehat{\gamma}_1^{LA} Z_i \quad (14)$$

An assumption in this “matching” procedure is that the YL and ENHAB samples share similar characteristics such as national representativeness.³ With this prediction we estimate the wage equation, and we analyze the gender wage gap as the theoretical model suggests: modeling wages as a function (basically) of latent cognitive and socioemotional abilities. The usual empirical approach is to model wages as a function of measured abilities (test scores) which leads to biased estimates. Thus, we exploit the calculated proxies for the latent abilities in the ENHAB to compare the usual approach with these results.

Finally, we use the Oaxaca–Blinder decomposition based on those estimated fixed effects as controls in the wage equation, and we estimate a theoretical model of log wages on latent ability as proposed by Heckman et al. (2006).

5.1 Gender wage gap and Oaxaca–Blinder decomposition

According to the model, measuring the wage gap based on test scores gives a wrong appreciation of the contributions of abilities and their returns to the gender wage gap. In particular, let us consider the following relation between wages and test scores:

$$\ln W = \gamma_Y X_Y + \gamma_Y^C C + \gamma_Y^S S + \epsilon_Y \quad (15)$$

Estimating this equation provides biased estimators of γ^C and γ^S . In this equation latent abilities are unobserved and considered in the error term s_Y . Since Cognitive (C) and Socioemotional (S) test scores are functions of latent abilities, they are correlated with the error term. Therefore, the estimated coefficients do not reproduce the effect of the latent abilities on wages. Once the gender wage gap is identified and the proxies of latent cognitive and socioemotional abilities are estimated and predicted we can apply one approach that allows us to evaluate the role of certain variables on the gender wage gap: the Oaxaca–Blinder decomposition.

The Oaxaca–Blinder decomposition is a method that aims to decompose differences in mean wages across two groups, in this case, between genders. The setting assumes a linear model that is separable in observable and unobservable characteristics:

$$Y_g = X\beta_g + \eta_g \text{ for } g = \text{male, female} \quad (16)$$

Thus, letting d be an indicator variable for group membership, y^d be the scalar outcome of interest for a member group d , X^d be a vector of observable

³ We should consider that the YL sample ignores children in the top 5% of the national income distribution.

characteristics (including a constant), $\widehat{\beta}^d$ be the column vector of coefficients from a linear regression of y^d on X^d , and letting overbars denote means, one can re-express different wages between different observable characteristics or differences in coefficients:

$$\overline{Y^1} - \overline{Y^0} = (\overline{X^1} - \overline{X^0}) \widehat{\beta}^1 + \overline{X^0} (\widehat{\beta}^1 - \widehat{\beta}^0) \quad (17)$$

where the first and second terms on the right-hand side of the equation represent the explained and unexplained components of the difference in mean outcomes, respectively. This is what we call a “two-fold decomposition”. An extension of this method is called the “three-fold decomposition” which includes a third term that interacts (simultaneous) differences in observable characteristics with coefficients:

$$\overline{Y^1} - \overline{Y^0} = (\overline{X^1} - \overline{X^0}) \widehat{\beta}^1 + \overline{X^0} (\widehat{\beta}^1 - \widehat{\beta}^0) + (\overline{X^1} - \overline{X^0}) (\widehat{\beta}^1 - \widehat{\beta}^0) \quad (18)$$

where the last term on the right-hand side of the equation represents the interaction.

6 Results

In this section, we compare the results of estimating the effect of cognitive and socioemotional abilities on wages by using measures of these skills (test scores) with those obtained by using two definitions of latent abilities. In each case we start by presenting the Mincer equation of log wages that controls for schooling and abilities. Then, we apply the Oaxaca–Blinder decomposition to estimate the effect of the abilities on the gender wage gap. Further, to disentangle this effect in each of the choices made by the individual before earning a certain wage we estimate a joint model of schooling, employment, occupational choice, and wages. To proceed in this manner, we apply the procedure explained previously to obtain proxies for cognitive and socioemotional abilities and present the results obtained in each of the four stages.

6.1 Wages and measured abilities

Considering the previous discussion on the issues of estimating the effect of abilities on wages, Table 3 shows the results of a basic Mincer equation under the naïve assumption that there is no correlation between measured skills and schooling. Column 1 of Table 3 shows that after controlling for work experience, place of residence, mother tongue, and birth order, an additional year of schooling leads to a 11.3% increase in earnings. In column 2, we present results that control for the parents’ schooling, as it may explain part of the correlation between earnings and schooling. As suspected, the point estimate drops from 0.113 to 0.0983. The inclusion of measures of cognitive and socioemotional abilities shows that cognitive ability and emotional stability lead to higher wages while agreeableness and consistency of effort reduce it.

Table 3 Mincer equation with measured abilities

	[1]	[2]	[3]
Schooling	0.1130*** (0.0083)	0.0983*** (0.0079)	0.0817*** (0.0121)
Experience	0.0266 (0.0166)	0.0329* (0.0162)	0.0340* (0.0174)
Experience 2	-0.0004 (0.0004)	-0.0004 (0.0004)	-0.0005 (0.0004)
Lives in Lima	0.0458 (0.0319)	0.0355 (0.0320)	0.0149 (0.0315)
Mother tongue (Spanish)	-0.0369 (0.1061)	-0.1091 (0.1129)	-0.1451 (0.1089)
Order of birth	0.0070 (0.0088)	0.0143 (0.0086)	0.0158* (0.0087)
Father's educational level		0.0088 (0.0054)	0.0061 (0.0046)
Mother's educational level		0.0167* (0.0081)	0.0161* (0.0079)
Goldberg, extraversion			0.0273 (0.0339)
Goldberg agreeableness (kindness)			-0.0485 (0.0299)
Goldberg, agreeableness (cooperation)			-0.0570 (0.0365)
Goldberg, conscientiousness (strong)			-0.0235 (0.0260)
Goldberg, emotional stability			0.0705*** (0.0213)
Goldberg, openness			0.0039 (0.0248)
Grit 2, consistency of interest			-0.0368 (0.0219)
Grit 2, persistence effort			0.0038 (0.0254)
Cognitive test score			0.0961*** (0.0239)
<i>N</i>	1080	1080	1080
<i>R</i> -squared	0.15	0.16	0.18

Robust standard errors in parentheses are clustered at the regional level. Sample is all people who are working at the moment of the interview and have available information on labor outcomes, test scores, and individual controls

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

6.1.1 Oaxaca–Blinder decomposition

In order to estimate the contribution of measured abilities on the gender wage gap we apply the Oaxaca–Blinder decomposition. The sample mean of log hourly wages is 1.430 for men and 1.167 for women and yields a statistically significant wage gap of 0.263. The wage gap can be attributed to differences in the predictors and in the coefficients. Nonetheless, while there would be a significant increase in women's hourly wages if they had the same characteristics (mean values of the regressors) as men, around 80% of the gender wage gap would be reduced if women shared the men's coefficients or returns, given their own characteristics. The gender wage gap as well as the endowment effect and the differences in coefficients are significant even after controlling for standard individual characteristics. Columns 1 and 2 of Table 4 illustrate the results obtained by means of the two-fold decomposition using a simple specification and after adding standard controls, respectively. Columns 3 and 4 present the results obtained by means of the three-fold decomposition.

Table 4 Oaxaca–Blinder decomposition with measured abilities

	[1]	[2]	[3]	[4]
Wage gap	0.263*** (0.061)	0.263*** (0.060)	0.263*** (0.061)	0.263*** (0.060)
Endowment	0.062*** (0.021)	0.046* (0.024)	0.071*** (0.026)	0.057** (0.028)
Return	0.200*** (0.058)	0.216*** (0.057)	0.208*** (0.062)	0.226*** (0.056)
<i>Endowment</i>				
Cognitive	0.063*** (0.020)	0.063*** (0.020)	0.072*** (0.025)	0.072*** (0.025)
Non-cognitive	-0.001 (0.002)	-0.000 (0.001)	-0.001 (0.002)	-0.000 (0.002)
<i>Return</i>				
Cognitive	-0.190 (0.179)	-0.205 (0.169)	-0.182 (0.171)	-0.196 (0.162)
Non-cognitive	-0.002 (0.144)	-0.037 (0.151)	-0.002 (0.144)	-0.037 (0.152)
<i>Observations</i>	1080	1080	1080	1080
<i>Controls</i>	No	Yes	No	Yes
<i>Interactions</i>	No	No	Yes	Yes

Robust standard errors in parentheses are clustered at the regional level. Sample is all people who are working at the moment of the interview and have available information on labor outcomes, test scores, and individual controls. Controls include age (and its square) and residence in Lima

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

The detailed decomposition shows that the explained part of the wage gap is mainly driven by the difference in endowments of cognitive ability between men and women. The results also show that (measured) non-cognitive or socioemotional ability⁴ does not make a significant individual contribution to the explained part of the gap.

As noted in Jones (1983) and Oaxaca and Ransom (1999), the interpretation of the detailed decomposition of the unexplained part of the gender gap (or returns) depends arbitrarily on the reference category of predictors. We abstain from interpreting the detailed decomposition of the unexplained part and focus on the contribution of the individual predictors to the explained part of the gender gap, which is unaffected by the choice of the base category.

6.1.2 Joint estimation: schooling, employment, occupation, and wages

The previous results describe the correlation between cognitive and socioemotional skills and wages but do not account for choices made by the individual before receiving a wage. To disentangle the effect of measured skills on each of these choices, we proceed with a joint estimation that considers sequential choices of schooling, employment, and occupation. The model follows an individual's line of choice. First, the individual is aware of his or her own level of abilities and chooses a schooling level using this information. After completing the chosen level of

⁴ From now on, we will work with the composite measure of GRIT as the representative test for measuring socioemotional ability. We chose to work with GRIT as the literature on socioemotional abilities highlight its importance, and because Díaz et al. (2012), who also use the ENHAB, find that it plays an important role on wage equations. Nonetheless, every estimation has also been performed with the measures of the rest of personality traits and arrives at similar results.

Table 5 Joint likelihood with measured abilities

	Employment	Hourly wages	Schooling	Occupational choice
Cognitive (female)	-0.000	0.007	0.047***	0.044***
Interaction w/cognitive	0.002	0.020	0.006	-0.007
Non-cognitive (females)	0.123**	0.844**	0.302***	0.115
Interaction w/socioemotional	0.229***	1.469***	-0.070	0.120
Observations	2407			

The interaction is (Male)*(Measured ability). Sample is all people who have available information on test scores and individual controls

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

schooling, the individual decides whether to enter or not the labor market. Once he or she decides to participate in the labor market, he or she chooses the occupation (white collar or blue collar) and, finally, receives a wage according to his or her previous decisions.

Table 5 shows the result of the maximum likelihood estimation of the joint model. The procedure requires the maximization of the joint likelihood of attaining a certain level of education, being employed, choosing a certain occupation, and earning a certain wage. Thus, the individual contribution to the likelihood is:

$$i_i = \underbrace{L_{si}(\theta_S|LA_i)}_{\text{Schooling}} \underbrace{L_{hi}(\theta_h|LA_i, s_i)}_{\text{Working}} \underbrace{L_{oi}(\theta_o|LA_i, s_i, h_i = 1)}_{\text{Occupation}} \underbrace{L_{wi}(\theta_W|LA_i, s_i, h_i = 1, o_i)}_{\text{Wages}} \tag{19}$$

In the above equation, $L_{si}(\theta_S|LA_i)$ is the conditional density schooling (having attended through secondary school), given the latent abilities LA_i . $L_{hi}(\theta_h|LA_i, s_i)$ is the likelihood of being employed, given the latent abilities and schooling s_i . The following term, $L_{oi}(\theta_o|LA_i, s_i, h_i = 1)$, is the likelihood of choosing a white collar job, given the latent abilities, schooling, and that the individual is employed ($h_i = 1$). Finally, $L_{wi}(\theta_W|LA_i, s_i, h_i = 1, o_i)$ represents the conditional distribution of wages, given the latent abilities, years of schooling, being employed, and occupational choice o_i .

Each column of Table 5 corresponds to each of the choices involved in the model. The results indicate that while measured cognitive skills seem to matter the most in determining the years of schooling and occupational choice, the measured socioemotional abilities gain relevance for wages and employment. In terms of inter-gender differences, men have higher returns to socioemotional abilities than women in terms of being employed and earning higher wages. Women have a higher return to cognitive abilities only in the choice of schooling. Nevertheless, these estimated contributions consider the measured abilities, which could be capturing the effect of other factors correlated with the outcome variables and measured abilities.

Table 6 First stage estimation (fixed effect model of measured ability)

	Non-cognitive ability (self esteem)	Cognitive ability
Caregiver's ability	0.039*** (0.013)	
Wealth index	0.027 (0.180)	-2.793 (1.854)
Stand. height for sge	-0.136*** (0.047)	-1.698*** (0.587)
Stand. body mass index	0.068*** (0.023)	-1.377*** (0.302)
Age in months	-0.001 (0.001)	0.793*** (0.013)
(Perceived) household status	0.343*** (0.035)	1.144*** (0.376)
Missed school due to illness	-0.139*** (0.048)	-1.416*** (0.462)
Log household consumption	0.087 (0.060)	-0.363 (0.553)
Observations	349	349
R-squared	0.035	0.820

Sample of children living in an urban area with available information on relevant variables. Clustered standard errors are at the community level

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

6.2 Wages and latent abilities

To properly estimate the contribution of abilities on the gender wage gap, we consider latent abilities. In the following subsections, we present the results of the proposed procedure for estimating latent cognitive and socioemotional abilities. Then, we estimate the previous models with the resulting proxies for the latent abilities instead of the measured abilities.

6.2.1 Estimating latent abilities and factor loadings

Table 6 shows the results of the first stage in columns 1 and 2 that correspond to cognitive ability (PPVT scores) and self-esteem, respectively. Each regression controls for the child's caregiver's measured self-esteem⁵ and the child's standardized height for age, standardized body mass index, age, and an indicator for having missed school for more than one week due to illness (-not so- exogenous variation in schooling) as well as the household's perceived status, wealth index, and log real consumption per capita. Standard errors are clustered at the community level.

In the case of self-esteem, changes in the caregiver's self-esteem and the child's height for age, body mass index, missed school for more than one week due to illness, and perceived household socioeconomic status (SES) are statistically significant. One interesting result is the large effect associated with SES. It indicates the status within social groups, elevated self-esteem should result from elevated SES (Rosenberg and Pearlin 1978). If an individual aspires to success in the form of social status and achieves these goals, elevated self-esteem should result (Twenge and Campbell 2002). This is especially relevant for adults whose SES reflects their own earned status (Rosenberg and Pearlin 1978). The research of socioemotional skills has a long way to go; however, that is beyond the scope of this paper. Furthermore,

⁵ No measure of the caregiver's cognitive ability was available in the dataset.

we include additional control variables in both estimations finding that the magnitude and level of significance of coefficients stayed the same as presented in Table 6.

Another interesting result is the negative association between BMI and cognitive ability. The BMI (body mass index) is an attempt to quantify the amount of tissue mass (muscle, fat, and bone) in an individual. Basically, it is a measure of overall healthiness. Because of poverty levels in Peru, malnourishment among children is common. Thus, a negative correlation between BMI and cognitive skills is rare. However, according to the data, a priori expectation of unhealthy children does not hold. The sample mean of BMI is around 18 points, which is a borderline level between underweight and healthy weight. An interesting fact is the number of overweight children ($BMI > 25$): 88 out of 349 children. Even though BMI does not measure body fat directly, a high BMI can be an indicator of high body fatness. Research shows that body fatness is negatively related with cognitive skills.

For instance, Baccouche et al. (2014) estimate the relation between BMI and cognitive performance in rugby players and find that BMI is negatively correlated with verbal proficiency. In the same vein, Smith et al. (2011) summarize diverse studies that point out that increased adiposity is associated with poor cognitive performance, especially in the executive functions in children, adolescents, and adults. Also, Basatemur et al. (2013) find that maternal pre-pregnancy BMI (which determines children's BMI) is negatively associated with children's cognitive performance, even after adjusting for multiple sociodemographic confounders. They also find that the relation appears to become stronger as children get older. Identification of the causes of this negative association requires a more thorough investigation. However, the literature seems to be shedding some light on the subject.

Table 7 shows the results of the second stage of our procedure. In this stage, we estimate the coefficients associated with the permanent characteristics that we use to predict latent ability using the ENHAB sample. Based on the literature and the

Table 7 Second stage estimation (latent ability on permanent characteristics)

	Non-cognitive ability (self esteem)	Cognitive ability
Sex (male)	0.011 (0.016)	0.984*** (0.360)
Order of birth	-0.019*** (0.003)	-1.000*** (0.155)
Father's schooling	0.033*** (0.006)	1.283*** (0.078)
Mother's schooling	0.015*** (0.002)	0.297** (0.132)
Mother tongue (Spanish)	-0.087*** (0.018)	-0.654 (0.442)
Constant	-0.424*** (0.068)	-55.256*** (2.003)
Observations	349	349
R-squared	0.052	0.186

Sample of children living in an urban area with available information on relevant variables. Clustered standard errors at the community level

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

Table 8 Third stage statistics (predicted latent abilities in both databases)

	Mean Female	Mean Male	Mean Difference	Whole sample Mean
Young lives sample				
Latent cognitive ability	-42.676	-40.8554	1.821*** (0.180)	-41.716
Latent self-esteem	-0.067	-0.036	0.031*** (0.005)	-0.050
Observations	165	184		349
ENHAB sample				
Latent cognitive ability	-47.990	-46.676	1.314*** (0.233)	-47.353
Latent self esteem	-0.241	-0.221	0.019*** (0.006)	-0.231
Observations	3844	3617		7461

Predicted fixed effects for the ENHAB sample were based on YL estimates for the urban subsample. Clustered standard errors at the regional level are in parentheses

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

availability of these variables on both the YL and EHAB datasets, we control for gender, order of birth, parents' educational level, and mother tongue. The literature on skill formation shows that the latent ability is innate and thus, should be affected by characteristics that are determined for the child up to its first three years of life. What we estimate as latent ability is actually the ability formed when the child was 12 years old, so one could expect, a priori, that variables that are fixed until that moment should be important in determining this latent ability. Covariates such as parents' educational level, gender, and the child's first language should be important, but not others such as characteristics of secondary education (which would also be endogenous). This is what motivates the reduced model. For the proxies for cognitive and non-cognitive latent abilities, all included controls are statistically significant for one measure or the other. Men have a higher endowment of cognitive ability, while the difference in self-esteem is not statistically significant after controlling for the order of birth, mother tongue, and parents' education. Lower ability, both cognitive and non-cognitive, is also related with a higher order of birth. Almost consistently, parents' education has a positive impact on both proxies of ability. Finally, the child's first language is an important determinant of self-esteem but not of cognitive ability.

The third stage of the procedure predicts the latent abilities in the ENAHB sample by using the coefficients and regressors displayed in Table 7. Since both the YL and ENHAB surveys are nationally representative, the matching procedure should be plausible. Table 8 shows the descriptive statistics of both predictions (for the YL and ENHAB samples) in the full sample and by gender. Due to limitations in our data, our predictions of cognitive and non-cognitive abilities are not able to explain a great part of the variation. However, both predictions share similar characteristics and directions in the gender differences that indicate our procedure is valid within this limitation.

Table 9 Mincer equation with latent abilities

	[1]	[2]	[3]	[4]
Schooling	0.1130*** (0.0083)	0.0983*** (0.0079)	0.0817*** (0.0121)	0.0953*** (0.0097)
Experience	0.0266 (0.0166)	0.0329* (0.0162)	0.0340* (0.0174)	0.0332* (0.0167)
Experience 2	-0.0004 (0.0004)	-0.0004 (0.0004)	-0.0005 (0.0004)	-0.0004 (0.0004)
Lives in Lima	0.0458 (0.0319)	0.0355 (0.0320)	0.0149 (0.0315)	0.0588** (0.0273)
Mother tongue (Spanish)	-0.0369 (0.1061)	-0.1091 (0.1129)	-0.1451 (0.1089)	
Order of birth	0.0070 (0.0088)	0.0143 (0.0086)	0.0158* (0.0087)	0.2372*** (0.0527)
Father's educational level		0.0088 (0.0054)	0.0061 (0.0046)	-0.2676*** (0.0630)
Mother's educational level		0.0167* (0.0081)	0.0161* (0.0079)	-0.0425** (0.0197)
Goldberg, extraversion			0.0273 (0.0339)	
Goldberg, agreeableness (kindness)			-0.0485 (0.0299)	
Goldberg, agreeableness (cooperation)			-0.0570 (0.0365)	
Goldberg, conscientiousness (strong)			-0.0235 (0.0260)	
Goldberg, emotional stability			0.0705*** (0.0213)	
Goldberg, openness			0.0039 (0.0248)	
Grit 2, consistency of interest			-0.0368 (0.0219)	
Grit 2, persistence effort			0.0038 (0.0254)	
Cognitive test score			0.0961*** (0.0239)	
Positive (latent) personality trait				0.0769* (0.0416)
Negative (latent) personality trait				-0.1017*** (0.0265)
Predicted (latent) non-cognitive ability				-0.5790 (1.381)
Predicted (latent) cognitive ability				0.2290*** (0.0599)
N	1080	1080	1080	1080
R-squared	0.15	0.16	0.18	0.19

Robust standard errors in parentheses are clustered at the regional level. Sample is all people who are working at the moment of the interview and have available information on labor outcomes, test scores, and individual controls

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

Finally, we estimate the effect of latent abilities on wages. Table 9 compares the results of the basic Mincer equation obtained by including the measured abilities (column 3) and those obtained by controlling, instead, for our predicted latent abilities⁶ (column 4). In this regression, we also control for personality traits by including two latent measures of Goldberg and GRIT items. A common latent factor arises from the principal component analysis of the measured personality traits that have a positive effect on wages (extraversion, emotional stability, openness and effort). We call this factor the latent positive trait. Similarly, we capture a latent common factor, the “latent negative trait”, from the Goldberg and GRIT items that have a negative relation with earnings (kindness, cooperation, conscientiousness, and consistency of interest).

Two results are worth highlighting. First, the return to schooling in column 4 is larger than that of column 3. This is consistent with our previous suspicion that measured abilities capture part of the effect of schooling on wages (the reason behind the drop in returns to schooling from column 2 to column 3). Second, the statistical significance of latent cognitive abilities. This significance shows that there is an effect of cognitive abilities on wages, but also that it now also captures the indirect effect of abilities through schooling now that we are able to control for both schooling and latent abilities. As expected, the latent negative and positive personality traits have significant effects on wages, and we find no significant effect of socioemotional skills on wages.

6.2.2 Oaxaca–Blinder decomposition

This subsection describes the results obtained after applying the Oaxaca–Blinder decomposition to the whole sample of the working age population⁷ but accounting for differences in latent abilities (the previously estimated proxies). As stated in the specification with measured abilities, a significant gender wage gap exists and, again, this gap is largely explained by returns (or the unexplained part of the wage gap). Regarding the explained part of the gap, in contrast to the results obtained in the previous subsection, we find that socioemotional skills also play an important role in explaining the wage gap.

Table 10 shows the results corresponding to the Oaxaca–Blinder decomposition for the ENHAB sample that accounts for differences in latent cognitive ability as well as latent self-esteem (as a proxy for the socioemotional latent ability) and also controls for latent personality traits. After applying the two-fold and three-fold approximations, the data support the fact that the gender wage gap is attributable to group differences in the coefficients and the predictors. In terms of differences in the endowment of abilities, Table 10 shows that the differences in cognitive and socioemotional abilities favor men regarding earnings. While the higher endowment of cognitive ability amongst men appears to reduce the gender wage gap, if women

⁶ We also analyzed the relation between latent abilities and wages using another measure of non-cognitive ability (self-efficacy) and obtained similar results.

⁷ The sample size is higher than in the O-B section with measured abilities because we also consider those with no information on measured abilities. We proceed in this way in order to exploit the variability in the available data as much as possible.

Table 10 Oaxaca–Blinder decomposition with latent abilities

	[1]	[2]	[3]	[4]
Wage gap	0.262*** (0.056)	0.262*** (0.053)	0.262*** (0.056)	0.262*** (0.053)
Endowment	0.033* (0.019)	0.028 (0.025)	0.049** (0.022)	0.054** (0.026)
Return	0.229*** (0.051)	0.234*** (0.048)	0.246*** (0.054)	0.264*** (0.045)
<i>Endowment</i>				
Cognitive (latent)	-0.115* (0.060)	-0.077 (0.078)	-0.093 (0.068)	-0.020 (0.099)
Non-cognitive (latent)	0.116** (0.056)	0.099 (0.065)	0.106* (0.060)	0.063 (0.076)
Positive personality trait (latent)	0.024** (0.010)	0.020** (0.009)	0.027** (0.012)	0.023** (0.011)
Negative personality trait (latent)	0.008 (0.007)	0.008 (0.007)	0.009 (0.009)	0.009 (0.008)
<i>Return</i>				
Cognitive (latent)	0.953 (1.842)	2.587 (1.886)	0.977 (1.890)	2.654 (1.934)
Non-cognitive (latent)	-0.106 (0.315)	-0.383 (0.312)	-0.118 (0.350)	-0.426 (0.346)
Positive personality trait (latent)	-0.007 (0.007)	-0.008 (0.006)	-0.003 (0.004)	-0.004 (0.005)
Negative personality trait (latent)	0.002 (0.003)	0.001 (0.003)	0.003 (0.005)	0.002 (0.004)
<i>Observations</i>	1115	1115	1115	1115
<i>Controls</i>	No	Yes	No	Yes
<i>Interactions</i>	No	No	Yes	Yes

Robust standard errors in parentheses are clustered at the regional level. Sample is all people who are working at the moment of the interview and have available information on labor outcomes, predicted latent abilities, and individual controls. Controls include age (and its square) and residence in Lima

*Indicates 10% significance level; **indicates 5% significance level; and ***indicates 1% significance level

had the same endowment of socioemotional ability as men, they would earn significantly higher wages. The difference in the endowments of the positive personality traits favors men, and also explains part of the wage gap.

6.2.3 Joint estimation: schooling, employment, occupation, and wages

Table 11 shows the results of the joint estimation that considers cognitive and socioemotional latent abilities. Interpreting the role of both abilities in each of the choices considered lead to interesting results. First, the non-cognitive ability is crucial for attaining higher levels of education and this is so for men and women. Cognitive abilities are not determinants for this choice. The other three choices must be interpreted together. We observe that self-esteem is important in determining occupational choice. Although the interaction term is not statistically significant, the direction of the coefficient points to an advantage for men regarding occupational choice. We could interpret that men earn higher wages because when employed their equilibrium assignment is towards occupations with higher rewards for cognitive skill. This, combined with the fact that men have higher cognitive skills, helps explain the gender wage gap.

In contrast with the results from the joint estimation that uses measured abilities, we can observe that inter-gender differences in cognitive and socioemotional abilities in favor of men drop when considering latent abilities. Moreover, returns to non-cognitive latent abilities gain significance for the occupational choice. This significance supports our idea that most differences attributable to abilities occur within occupational choice.

Table 11 Joint likelihood with latent abilities

	Employment	Wages	Schooling (tertiary)	Occupational choice
Cognitive (female)	-0.040**	-0.382***	0.010	-0.040
Interaction w/cognitive	0.026	0.211	-0.027	-0.030
Non-cognitive (females)	1.026*	10.621**	2.248**	3.335***
Interaction w/socioemotional	-1.130	-8.841	0.857	1.174
Observations	7491			

The interaction is (Male)*(Latent Ability)

*Indicates 10% significance level; **indicates 5% significance level; ***indicates 1% significance level

7 Conclusions

This study presents evidence on the role of cognitive and socioemotional skills in closing the gender wage gap. In a first attempt to estimate their effect on wages, we followed the basic empirical approach of modeling wages in terms of measured ability. Second, we applied a procedure that estimated a model in terms of latent ability. We based the model on the setting proposed by Heckman et al. (2006). While these authors identify latent abilities based on dependence on different test scores for the same time period, we use variation over time for the same test score. This is possible due to the availability of panel data information on measures of cognitive and socioemotional skills. In addition, we estimate a joint model of schooling, employment, occupational choice, and wages in order to disentangle the effects of the latent abilities in the gender wage gap throughout an individual's choices previous to earning a certain wage. Our main contribution is analyzing the role of latent socioemotional and cognitive abilities to the gender wage gap in a developing country by estimating and accounting for proxies of the latent abilities and disentangling the effect of these abilities by means of a joint model of schooling, employment, occupational choice, and wages.

There is a significant gender wage gap in Peru. Estimations with measured abilities confirm the empirical literature regarding endogeneity issues that result from using test scores as measures of ability. The Oaxaca–Blinder decomposition in a model with measured abilities shows the significant inter-gender differences in the endowment of cognitive skills but no relevant differences in terms of socioemotional abilities (in endowment or returns). Estimating the joint model shows that differences in socioemotional abilities between men and women are important but only for choices prior to wage determination. Cognitive skills are relevant in determining years of schooling and occupational choice and measure the socioemotional ability for wages and employment.

Acknowledgements We want to thank Sergio Urzúa, Omar Arias, Marcel Fafchamps, five anonymous referees, and seminar participants in the “Conference on Skills, Education and Labor Market Outcomes” at the University of Maryland, “Inequalities in Children’s Outcomes in Developing Countries Conference” at Oxford University, and the Lacea and Peruvian Economic Association for helpful comments and discussions.

References

- Altonji, J. G., & Blank, R. (1999). Race and gender in the labor market. *Handbook of Labor Economics*, 3c, 3144–4259.
- Baccouche M. A., Arous I., Sellami H., & Elloumi A. (2014). Association between body mass index and cognitive performance in rugby players. *International Journal of Scientific and Research Publications*, 4(6).
- Basatemur E., Gardiner J., Williams C., Melhuish E., Barnes J., & Sutcliffe A. (2013). Maternal pre-pregnancy bmi and child cognition: a longitudinal cohort study pediatrics. *Pediatrics*, 131(1), 56–63.
- Behrman, J. R., Ross, D. R., & Sabot, R. H. (2008). Improving the quality versus increasing the quantity of schooling: evidence for rural Pakistan. *Journal of Development Economics*, 85(1), 94.
- Behrman J.R., Hoddinott J.F., Maluccio J.A., & Martorell R. (2011) *Brains versus brawn: labor market returns to intellectual and physical human capital in a poor developing country*. University of Pennsylvania
- Bell, E. C., Willson, M. C., Wilman, A. H., Dave, S., & Silverstone, P. H. (2006). Males and females differ in brain activation during cognitive tasks. *Neuroimage*, 30(2), 529–538.
- Bertrand, M. (2011). New perspectives on gender. In O. C. Ashenfelter, & D. Card (Eds.), *Handbook of labor economics* (Vol. 4B, pp. 1543–1590). Amsterdam: North-Holland. [https://doi.org/10.1016/S0169-7218\(11\)02415-4](https://doi.org/10.1016/S0169-7218(11)02415-4)
- Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), 789–865.
- Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: educational reform and the contradictions of economic life*. New York, NY: Basic Books.
- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31–47.
- Cunha, F., & Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738–782.
- Cunha F., Heckman J. J., Lochner L., & Masterov D. V. (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek, & F. Welch (Eds) *Handbook of the economics of education*. Edition 1, (Vol. 1, Chapter 12, pp. 697–812). Elsevier.
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 883–931.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4), 1593–1640. <https://doi.org/10.1093/qje/qjx022>.
- Díaz J.J. Arias O., & Vera Tudela D. (2012). Does perseverance pay as much as being smart?: The returns to cognitive and non-cognitive skills in urban Peru, mimeo.
- Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.
- Edwards, R. (1976). Individual traits and organizational incentives: what makes a “good” worker? *Journal of Human Resources*, 11(1), 51–68.
- Fortin, N. M. (2008). The gender wage gap among young adults in the United States: the importance of money versus people. *Journal of Human Resources*, 43(4), 884–918.
- Goldberg, L. R. (1990). An alternative “description of personality”: the big-five factor structure. *Journal of Personality and Social Psychology*, 59, 1216–1229.
- Groger, J., & Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources*, 30(2), 280–310.
- Grove, W. A., Hussey, A., & Jetter, M. (2011). The gender pay gap beyond human capital: Heterogeneity in noncognitive skills and in labor market tastes. *Journal of Human Resources*, 46(4), 827–8.
- Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1-2), 39–98.
- Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607–668.
- Heckman, J. J., Stixrud, J., & Urzúa, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269(5220), 41–45.
- Jones (1983). On decomposing the wage gap: a critical comment on Blinder’s method. *The Journal of Human Resources*, 18, 126–130.

- Klein, R., Spady, R., & Weiss, A. (1991). Factors affecting the output and quit propensities of production workers. *Review of Economic Studies*, 58(2), 929–954.
- Landes, L. (1977). Sex differences in wages and employment: a test of the specific capital hypothesis. *Economic Inquiry*, 15(4), 523–538. <https://doi.org/10.1111/j.1465-7295.1977.tb01115.x>. Retrieved from.
- Manning, A., & Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, 118(530), 983–1024. <https://doi.org/10.1111/j.1468-0297.2008.02158.x>.
- Mincer, J., & Ofek, H. (1982). Interrupted work careers: depreciation and restoration of human capital. *Journal of Human Resources*, 17(1), 3–24. <https://doi.org/10.2307/145520>.
- Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, 77(2), 251–266.
- Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4), 547–568.
- Neal, D. A., & Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5), 869–895.
- Oaxaca, R. L., & Ransom, M. R. (1999). Identification in detailed wage decompositions. *The Review of Economics and Statistics*, 81, 154–157.
- Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1), 123–144.
- Ragan, J. F., & Smith, S. P. (1981). The impact of differences in turnover rates on male/female pay differentials. *Journal of Human Resources*, 16(3), 343–365. <https://doi.org/10.2307/145625>.
- Rosenberg, M., & Pearlin, L. I. (1978). Social class and self-esteem among children and adults. *American Journal of Sociology*, 84, 53–77.
- Smith, E., Hay, P., Campbell, L., & Trollor, J. N. (2011). A review of the association between obesity and cognitive function across the lifespan: implications for novel approaches to prevention and treatment. *Obesity Reviews*, 12(9), 740–755.
- Strand, S. (2003). Sex differences in cognitive abilities test scores: a national picture. *Space*, 98, 1–28.
- Twenge, J. M., & Campbell, W. K. (2002). Self-esteem and socioeconomic status: a meta-analytic review. *Personality and Social Psychology Review*, 6(1), 59–71.
- Upadhyay, N., & Gurugain, S. (2014). Comparison of cognitive functions between male and female medical students: a pilot study. *Journal of clinical and diagnostic research*, 8(6), BC12.
- Urzúa S., Bravo D., & Sanhueza C. (2009). Ability, schooling choices and gender labor market discrimination: evidence for Chile. In: H. Ñopo, A. Chong, A. Moro (Eds.), *Discrimination in Latin America: an economic perspective*. Inter-American Development Bank.
- Urzúa, S. (2008). Racial labor market gaps: the role of abilities and schooling choices. *Journal of Human Resources*, 43(4), 919–971. <https://doi.org/10.3368/jhr.43.4.919>.
- Weinberger, C. J. (1999). Mathematical college majors and the gender gap in wages. *Industrial Relations*, 38, 407–413.
- Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. *Review of Economics and Statistics*, 96(5), 849–861. https://doi.org/10.1162/REST_a_00449.