118

Working Paper

Young Lives

An International Study of Childhood Poverty

Cognitive and Non-cognitive Skills and Wages

The Role of Latent Abilities on the Gender Wage Gap in Peru

Pablo Lavado, Luciana Velarde and Gustavo Yamada

www.younglives.org.uk

DECEMBER 2013

Working Paper 118

Young Lives
An International Study of Childhood Poverty

Cognitive and Non-cognitive Skills and Wages

The Role of Latent Abilities on the Gender Wage Gap in Peru

Pablo Lavado, Luciana Velarde and Gustavo Yamada

This paper was presented at a conference on Inequalities in Children's Outcomes in Developing Countries hosted by Young Lives at St Anne's College, Oxford on 8-9 July 2013.

http://www.younglives.org.uk/news/news/children-inequalities-younglives-conference-2013

The data used come from Young Lives, a longitudinal study of childhood poverty that is tracking the lives of 12,000 children in Ethiopia, India (in Andhra Pradesh), Peru and Vietnam over a 15-year period. **www.younglives.org.uk**

Young Lives is funded from 2001 to 2017 by UK aid from the Department of International Development and co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014.

Cognitive and Non-cognitive Skills and Wages: The Role of Latent Abilities on the Gender Wage Gap in Peru

Pablo Lavado, Luciana Velarde and Gustavo Yamada

First published by Young Lives in December 2013

ISBN: 978-1-909403-31-4

A catalogue record for this publication is available from the British Library.

Printed on FSC-certified paper from traceable and sustainable sources.

Funded by

UKaid from the British people

Ministry of Foreign Affairs of the Netherlands

Young Lives, Oxford Department of International Development (ODID), University of Oxford, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, UK

Tel: +44 (0)1865 281751 • E-mail: younglives@younglives.org.uk

Contents

Abstract
Recently there has been growing interest in the relationship between cognitive and non cognitive abilities and labor market outcomes. A large literature provides evidence on the positive connection between cognitive test scores and higher wages. Fewer and newer papers have explored the correlation between non cognitive test scores and wages. However, attention is focused on developed countries. Test scores suffer two limitations. First, they can be considered outcomes of the schooling level and latent (unobserved) cognitive and non cognitive abilities. Second, they are potentially measured with error. The main objective of this paper 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 Young Lives Study (YL) and the Peruvian Skills and Labor Market Survey (ENHAB). Young Lives has panel data information on test scores and ENHAB has information on test scores and wages. Results show that even though when accounting for measured abilities differences in non cognitive abilities seem irrelevant, when accounting for diferences in actual latent ability non cognitive 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 not only in wage profiles, but in schooling, employment and occupation decisions.

Acknowledgements
We would like to thank comments from Sergio Urzúa and seminar participants in the "Conference on Skills, Education and Labor Market Outcomes" at University of Maryland and in the "Inequalities in Children's Outcomes in Developing Countries Conference" at Oxford University, Universidad del Pacífico and Central Bank of Peru.

About the Authors
Gustavo Yamada is full profesor at Universidad del Pacífico and researcher at Centro de Investigación de la Universidad del Pacífico in Lima, Peru. He has a PhD in Economics (Columbia University). His areas of interest are Labor Economics, Economics of Education and Impact Evaluation.
Pablo Lavado is assistant professor at Universidad del Pacífico and researcher at Centro de Investigación de la Universidad del Pacífico in Lima, Peru. He has a PhD in Economics (CEMFI). His areas of interest are Labor Economics, Economics of Education and Impact Evaluation.
Luciana Velarde is currently working at Ministry of Development and Social Inclusion of Peru. She has a MA in Economics (Universidad del Pacífico). Her areas of interest are Economics of Education, Poverty and Development.

1 Introduction

Differences in cognitive skills are strongly related to differences in wages between females and males. Specifically, males have higher cognitive test scores than females on average and this contributes to higher wages for the former than for the latter. 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 bigger gain in terms of wage for a male than for a female, ceteris paribus.

Recently, literature has focused on the relationship between non cognitive skills and productivity. First, regarding the relationship between non cognitive test scores and labor market outcomes, the main finding is that there exists a positive connection between wages and certain non cognitive skills. Second, regarding how non cognitive skills are formed, it has been proposed that test scores are bad proxies of abilities due to measurement errors and endogeneity with schooling. The main important features that drive wages are latent abilities. However, to the best of our knowledge, there are few studies about the contribution of differences in latent non cognitive skills to the gender wage gap. Moreover, this question has not been addressed for the developing world and in Latin American Countries in particular.

Based on the latent ability model proposed by Heckman et al. (2006) we propose a method to estimate latent cognitive and non cognitive abilities. In contrast to Heckman et al. (2006) my identification strategy is based on panel data information 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 used them in a Oaxaca-Blinder decomposition in order 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 in order to disentangle the effect of intergender differences in ability in each of this choices. Since the Young Lives database lacks information on wages, we estimate latent abilities as linear combinations of characteristics common to both YL and ENHAB. The ENAHB is a recent survey in Peru which gathers data on cognitive and nongocnitive test scores, individidual's characteristics, educational trajectory and wages. we predict latent abilities in this data base and, afterwards, use these predictions in a Oaxaca-Blinder decomposition.

The main objective of this paper is twofold. First, we want to document differences in the distribution of non cognitive skills by gender. Second, we seek to estimate the contribution of non cognitive skills to the gender wage gap . Using a recent survey of the working age (14-50) urban population in Peru we apply the Oaxaca-Blinder decomposition for the gender wage gap. In an attempt to disentangle the forces behind the returns to non cognitive abilities, we will present a standard production function model where schooling, employment, occupational choice and earnings are a function of cognitive and non cognitive abilities and try to identify the parameters with the available information.

The main contribution of this paper is to analyze the role of latent non cognitive and cognitive abilities to the gender wage gap in a developing country by estimating and accounting for proxies of latent abilities. In order to achieve this we inspire my approach with that of Heckman et al. (2006), but exploiting panel data information to achieve identification.

Preliminary results show that there is a significant gender wage gap in Peru. In fact, in a model with measured abilities, we find significant intergender differences in the endowment of cognitive skills but no relevant differences in terms of non cognitive abilities (in endowment or returns). Estimating the joint model evidences that differences in non cognitive abilities between men and women are important but on choices prior to wage determination. Cognitive skills seem to be relevant in determining years of schooling and occupational choice and measured non cognitive ability for wages and employment. Applying our proposed estimation procedure revealed that actual latent ability turned out to be highly statistically significant for mean wages as well as accounting for inter-gender differences. In particular differences in the endowment of non cognitive abilities contribute (negatively) to the gender wage gap. No significant intergender differences in the returns to cognitive abilities were found. Moreover, the estimation of the joint model sheds light towards the fact that the observed gender wage gap is mainly attributed to differences within occupational choice. Cognitive and non cognitive abilities are valued differently for men and women in terms of schooling, employment and wages, but basically men seem to earn higher wages because their equilibrium assignation is towards occupations with higher rewards to cognitive skills, which 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. Section 4 describes the data and sample. Section 5 develops our econometric implementation for estimating latent abilities. Section 6 presents results. Section 7 concludes and proposes issues for further research.

2 Literature Review

For decades researchers have focused on studying the relationship between test scores and labor market outcomes. These studies are mostly related to cognitive test scores. Murnane, Willett and Levy (1995) assess the role of mathematics skills of graduating high school senior on their wages at age 24 and found a positive and increasing effect of cognitive skills on wages (specially in years closer to graduation). In a more recent study, Cunha, Heckman, Lochner and Masterov (2005) state that cognitive ability seems to affect the likelihood of acquiring higher levels of education and advanced training as well as the economic returns to these activities.

Attention is growing towards non cognitive abilities and their relationship with labor market outcomes. Early work by Bowles and Gintis (1976), Edwards (1976) and Klein, Spady and Weiss (1991) show that non cognitive skills such as dependence and persistence are highly valued by employers. Recent studies such as that of Heckman, Stixrud and Urzua (2006) also support this fact with evidence of a positive relation between results in non cognitive test scores and labor market outcomes.

Differences in skills have heterogeneous effects on labor market outcomes. Specifically, literature has focused on explaining differences in wages betwen men and women due to differences in cognitive skills (Neal and Johnson, 1996; Ritter and Taylor, 2011). However few studies address the contribution of non cognitive skills to differences in wages between males and females.

For instance, Fortin (2008) investigates the impact of traits such as self-esteem, external locus of control, the importance of money/work and the importance of people/family on the gender wage gap. Using two single-cohort longitudinal surveys, the NLS72 and the NELS88, she finds that non cognitive factors account for a small but not trivial part -about 2 log points-of the gender wage gap of workers in their early thirties. In particular, regarding the importance of "money/work" and "people/family" rather than self-esteem and confidence, i.e. men tend to be more ambitious and value money more and women tend to choose altruistic jobs.

Grove et al. (2011) explore whether non cognitive variables can explain more of the wage gap in MBA professionals than human capital variables. Using a longitudinal survey of individuals who registered for the Graduate Management Admission Test (GMAT) between 1990 and 1998 they find that that 82% of the gender wage gap is explained by non cognitive skills and preferences regarding family, career and jobs which on average pay lower wages. Women in the sample reveal their preference for altruistic jobs whereas men placed more importance on wealth.

Cobb-Clark and Tan (2009) examine whether men's and women's non cognitive skills influence their occupational attainment and whether this contributes to the disparity in their relative wages. Using the Household, Income and labor Dynamics in Australia (HILDA) and focusing on employees aged 25 to 65 years for 2001 through 2006 they find that men's and women's non cognitive skills have a substantial effect on occupational attainment. Nonetheless, while the overall gender wage gap is 0.143 log points, 96.6% is due to disparity in the wages of men and women employed in the same activity. Thus, the most substantial component of the gender wage gap occurs within occupations and remains largely unexplained.

Most of the research regarding the role of cognitive and non cognitive test scores on labor market outcomes has been focused in developed countries. To the best of our knowledge only few address this issue in developing countries, in particular in Latin American Countries (LAC).

Bassi and Galiani (2009) use a survey with nation-wide data for young adults aged 25 to 30 in order to explore the role of cognitive and non cognitive test scores on log earnings. They find significant coefficients of both types of test scores and that these are smaller after controlling for education. This is likely related to the fact that measured skills (as opposed to latent skills) are affected by schooling, generating biased coefficients for test scores in regressions that don't control for schooling.

Díaz et al. (2012) estimate the returns to education, cognitive and non cognitive skills in Peru using the ENHAB on a sample of working age population and applying an instrumental variable approach in order to address issues regarding endogeneity of schooling. They find that schooling, cognitive and non cognitive skills are valued in the Peruvian labor market. In particular, one standard deviation increase on years of schooling generates an increase of 15% on earnings, while a change in cognitive skills and non cognitive skills of a similar magnitude generate a 9% and 5% to 8% increase on earnings, respectively.

Urzúa et al. (2009) go a step further into analyzing gender labor market discrimination in Chile using a data set containing information on labor market outcomes, schooling attainment, schooling performance and others related to individual's family environment. They follow previous papers that estimate labor market models with multiple sources of unobserved heterogeneity through cognitive and non cognitive abilities. Nonetheless, due to data limitations, they consider only one underlying source of unobserved heterogeneity as a combination of cognitive and non cognitive abilities. Their results suggest the existence of gender gaps in labor market variables such as experience, employment, hours worked and hourly wages that cannot be explained by observable or unobservable characteristics or underlying selecion mechanisms generating endogeneity. Nonetheless they find that results depend on the schooling level analyzed, in particular, women seem to be discriminated against in the labor market but among less educated groups. As far as we know, this is the only paper that addresses the role of cognitive and non cognitive (latent) abilities (accounting for the endogeneity of schooling) in the gender wage gap in a developing country.

The problem of using test scores is that they do not reflect real (latent) abilities because they are measured with error Heckman et al. (2006). Therefore, using these test scores in wage and schooling regressions is troublesome. Conditioning on schooling, both cognitive and non cognitive tests predict wages. However, schooling is a choice variable and thus the endogeneity of schooling must be addressed. Omiting 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 non cognitive) tests (for example, IQ) and achievement tests. Although IQ is well set by age 8, achievement tests have been demonstrated to be quite malleable and increasing with schooling. This creates a reverse causality problem.

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. They find that the effects of schooling on test scores are roughly linear across schooling levels and are larger for lower ability levels. Schooling increases the AFQT score on average 2-4 pp. They contribute in estimating the impact of schooling on measured test scores at various quantiles of the latent ability distribution. They present evidence that the measure of IQ used by Herrnstein and Murray is strongly affected by schooling. They use a model of test scores as function of latent ability (and other determinants) and schooling as function of latent ability (and other determinants). They account for ceiling effects (on easy tests perfect scores are achieved by children with different ability levels) and endogeneity of schooling (choice of date of entry into schooling and of final schooling level).

Helmers and Patnam (2011) investigate the determinants of children's cognitive and non cognitive test scores in Andhra Pradesh, India using a rich database for two cohorts aged between one and twelve. Exploiting panel data information, they estimate a Linear Structural Relations (LISREL) model which allows for the estimation of latent cognitive and non cognitive skill levels and parental investment and allows to link these variables to observed child, parental and household characteristics. They build on Cunha and Heckman (2007) in an effort to examine the dynamics of both cognitive and non cognitive scores as well as their relationship over time.They find evidence of self-productivity for cognitive skills and cross-productivity effects from cognitive on non cognitive skills. They focus their research on exploring the determinants behind the formation process of both skills. They find evidence in favor of the importance of parental investment and child health at age one (parental care during pregnancy and early childhood).

Thus, the main contribution of our analisis is exploring the role that latent cognitive and non cognitive abilities and their returns in terms of wages play in explaining the gender wage gap in a developing country. In this vein, one of our main objectives is to estimate the latent abilities that are unobserved for the econometrician but permanent over time.

3 Model

The model is based on Heckman et al. (2006), Cunha et al. (2010) and Cunha and Heckman (2008). Latent cognitive and non cognitive 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, she will earn a wage.

3.1 The Model for Wages

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

$$LnW_i = \beta_Y X_Y + \alpha_Y^C f^C + \alpha_Y^N f^N + e_Y$$

where X_Y is a vector of observed controls, β_Y is the vector of returns, α_Y^C and α_Y^N are the latent cognitive and non cognitive abilities, respectively and e_Y represents and idiosyncratic error term independent of all other factors. We assume that the prices are the same for workers of different schooling categories. We also assume that the returns of latent abilities are constant between schooling levels.[1]

The identification strategy is somewhat similar as that in Heckman et al. (2006). We restrict that latent cognitive ability only affects cognitive measures and latent non cognitive ability only affects non cognitive measures.The model of the cognitive measure is:

$$C = \beta_C X_C + \alpha_C f^C + e_C$$

Likewise, the model of the non cognitive measure:

$$N = \beta_N X_N + \alpha_N f^N + e_N$$

Our assumptions imply that conditional on X variables, the dependence across time of measurements come from f^N and f^C.

3.2 The Model for Schooling

Each individual chooses the level of schooling that maximizes her lifetime expected benefit. Following a linear-in-the-parameters specification, and letting I_s represent the net benefit associated with schooling level s:

$$I_s = \beta_S X_S + \alpha_s^C f^C + \alpha_s^N f^N + e_s$$

where s is the schooling level chosen by the individual among $\bar{S}$ possibilities, X_s is a vector of observed variables affecting schooling, β_s is its associated vector of parameters, α_s^C and α_s^N are the factor loadings associated with cognitive and non cognitive latent abilities, respectively, and e_s represents an idiosyncratic component assumed to be independent of f^N, f^C and X_s. The error terms for each schooling level are mutually independent.

The observed schooling level corresponds to:

$$D_s = argmax_{s\epsilon 1,\ldots,\bar{S}}[I_s]$$

We consider two educational levels: (i) complete secondary education or higher, and (ii) up to incomplete secondary education. Thus, we will employ an indicator variable $D_s = 1(I_s > 0)$ is an indicator of choice of attaining complete secondary education or a higher educational level.

[1] Allowing for different returns by schooling level is left for further research.

3.3 The Model for Employment

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

$$I_E = \beta_E X_E + \alpha_E^C f^C + \alpha_E^N f^N + e_E$$

where β_E, X_E, α_E^C,α_E^N 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 1 if the individual is employed and 0 otherwise. The error term is orthogonal to the control variables.

3.4 The Model for Occupational Choice

Let I_0 denote utility associated with choosing a white collar occupation (where the alternative is a blue collar occupation). we postulate the following linear model for I_0:

$$I_0 = \beta_0 X_0 + \alpha_E^C f^C + \alpha_E^N f^N + e_0$$

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

Both cognitive and factors are known by each individual but not for the econometrician and they are fixed by the time the individual makes her labor market choices. Controlling for this dependence is equivalent to controlling for the endogeneity in the model. The main problem is that latent abilities are unobserved for the econometrician. Wage equations usually are functions of measured abilities or test scores. However these test scores 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. This problem is even worse when exploring the gender wage gap and estimating by gender wage equations as functions of abilities.

In order to deal with this endogeneity problem, Heckman et al. (2006) estimate the distributions of latent abilities relying on having at least three measurements. In contrast, our identification strategy relies on having panel data information on certain 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 measured abilities, schooling, labor force participation, occupational choices and wages. Unfortunately, this database is not available in developing countries. We propose an empirical methodology which exploits two datasets. The first one is the Young Lives Study database for Peru. The Young Lives Study 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 was collected on 20 sites of 14 regions and represents 95% of the peruvian children 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), anthropometric measures and 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 focus on the Older Cohort which, for Peru, comprises around 700 children that were 8 years old by the beginning of the study (born in 1994-5). we work with the subsample of children with available information on items related to cognitive and non cognitive abilities as well as individual characteristics for Rounds 2 and 3.[2] Finally, we work with the subsample of children living in urban areas. The final sample comprises 349 individuals.

This subsample is evenly distributed among boys and girls (165 and 184, respectively), with an average age of 149 months and have a mean of 6 years of schooling in Round 2. Table I presents descriptive statistics on the main variables of interest from both Rounds as well as information on the child's mother tongue and parents' educational level from Round 1 (what we call "permanent characteristics"). Some facts worth highlighting are that in both rounds, while girls score below average in items related to cognitive ability, boys do so in self efficacy items (results are mixed for self-esteem between rounds). Most household characteristics and family background appear to be similar between genders. Caregivers' measured non cognitive abilities differ between childs' gender; boys' cargeviers show 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 appear between both rounds of the survey, fact that will be helpful for our identification strategy.The measures used to represent non cognitive abilities were built based on respondents' degree of agreement or disagreement with a number of statements related to psychosocial traits such as self-esteem and self-efficacy.[3].

[2]Information regarding non cognitive abilities wasn't collected during the first round for the Older Cohort or in the later Rounds for the Younger Cohort.

[3]For self-esteem, the statements explored in the Young Lives survey focus on positive and negative dimensions of pride and shame based on the Rosenberg Self-Esteem Scale, focused on dimensions of children's living circumstances. For self-efficacy we focused on 5 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 is measured on a 4-point Likert scale that ranges from strong agreement to strong disagreement. we constructed two indices (one for each trait) as the average score of these items and used the standarized indices for our estimations.

Table I: Descriptive Statistics: Young Lives Sample

	Round 2				Round 3			
	Whole Sample		**Female**	**Male**	**Whole Sample**		**Female**	**Male**
	Mean	SD			Mean	SD		
Cognitive Ability (PPVT raw score)	76.920	13.840	75.964	77.777	101.083	14.600	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 measure (standarized)	0.144	0.997	-0.012	0.284^{***}	0.145	0.925	0.136	0.154
Caregiver's Self-Esteem measure (standarized)	0.046	1.003	0.058	0.036	0.044	1.028	0.130	-0.033
Height-for-age (standarized)	-1.302	1.049	-1.337	-1.271	-1.306	0.863	-1.502	-1.129^{***}
Body Mass Index (standarized)	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 (years)	6.143	0.895	6.170	6.120	9.006	1.101	9.067	8.951
Missed school $>$ 1 week due to illness	0.054	0.227	0.055	0.054	0.069	0.235	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 per capita	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**

NOTE: Âž indicates 15% significance level;* indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level of the mean test between males and females.

Table II: Descriptive Statistics for the ENHAB sample

	N	Mean	SD	Females	Males
Cognitive Ability (PPVT raw score)	2421	42.001	14.932	40.103	45.039***
Consistency / Interest	2421	-0.004	0.998	0.036	−0.069**
Persistance / Effort	2421	0.001	1.001	-0.046	0.076***
Grit (Standardized)	2421	-0.002	0.997	-0.009	0.008
Extraversion	2421	0.002	1.000	-0.047	0.082***
Kindness	2421	0.002	1.006	0.039	−0.058**
Cooperation	2421	0.006	0.992	0.051	−0.065***
Conscientiousness	2420	0.000	1.001	0.056	−0.092***
Emotional Stability	2415	0.013	0.997	-0.072	0.149***
Openness	2415	-0.004	0.999	-0.059	0.085***
Log Hourly Wages	4063	1.316	0.815	1.166	1.408***
Monthly Earnings	4063	972.155	1284.643	738.700	1116.299***
Hourly Earnings	4063	5.264	6.550	4.638	5.651***
Weekly work hours	4063	51.127	18.482	46.339	54.084***
Experience	4063	25.430	13.582	25.314	25.502
Age	7499	33.514	15.282	33.305	33.736
Schooling	7457	10.701	3.373	10.524	10.890***
Mother tongue (spanish)	7499	1.008	0.142	1.007	1.009
Father's educational Level	7499	4.955	2.373	4.947	4.963
Mother's educational level	7499	4.287	2.333	4.253	4.324

Note: * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level of the mean test between males and females.

The second database corresponds to a novel household survey collected by the World Bank in 2010 that not only contains information on wages and individual characteristics, but also on measured cognitive and non cognitive abilities for a sample of currently employed working age individuales. The Skills and Employability National Survey (ENHAB) is a nationally representative household survey that comprises information on urban areas (2600 households in cities with more than 70 000 inhabitants) of the country. Data collected contain information on household living conditions, demographic information, academic achievement, employment/earnings and novel information on (i) cognitive and non cognitive test scores, (ii) schooling trajectories, (iii) early labor market participation, and (iv) family characteristics. Measured abilities were assessed by means of cognitive tests evaluating numerical and problem solving skills, working memory, verbal fluency and receptive language, and non cognitive abilities according to GRIT scales (Duckworth et al., 2007) and the Big Five Personality Factors (Goldberg, 1990). For this analysis we focus on 7 of these measures, the standarized values of each the big-five factors (emotional stability, extraversion, agreableness/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 include personal educational background, family characteristics and socio-economic 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 will work with three subsamples that resemble each other in most statistics: (i) individuals with available information on measured abilities (test scores), N=2421; (ii) individuals with positive earnings, N=4063; (iii) individuals with available information on relevant individual characteristics, N=7499. In general terms, individuals in the data set are evenly distributed among men and women, have a mean age of 33 years, monthly earnings of around 1000 soles in constant peruvian currency to year 2010 (around 350 USD), work an average of 51 hours a week and have on average complete secondary education. Table II show some other relevant descriptive statistics for the three subsamples and the difference in each between men and women. Some facts worth highlighting are that men earn higher earnings (monthly and hourly), work longer hours and have higher levels of measured cognitive abilities than women. However, results are mixed regardin non cognitive skills. While women appear to be more consistent, kind, cooperative and conscientious,men appear to be more persistent, extravertive, 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:

$$LnW_i = \alpha + \gamma S_i + \beta_A A_i + \mu_i \tag{1}$$

where LnW_i are log earnings, S_i represents years of schooling and A_i is ability: Cognitive (C) and Non-cognitive (N). The main problem of this equation is that A_i is unobserved by the econometrician. Thus, if schooling is correlated with ability, and ability is ommitted, 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 of these abilities:

$$LnW_i = \alpha + \gamma S_i + \beta_T T_i + v_i \tag{2}$$

where T_i are standarized test scores for measured cognitive and/or non cognitive 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 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. Morover, since A is still ommitted, γ and β_T are overestimated.

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

First, we use time variation (from Round 2 to Round 3) in measured cognitive and non cognitive test scores and years of schooling among children in the YL sample to recover the (unobserved) fixed effects. In particular, we try to explain variation in three measures of ability, two non cognitive abilities (self-esteem and self efficacy) and one cognitive ability (Peabody Picture Vocabulary Test scores). The identification procedure requires controlling for characteristics that may have varied between the ages of 12 and 15, and that may explain variation in measured abilities during that period. In this way we will be able to explain changes in 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.

$$\Delta MA_{it} = \gamma_X \Delta X_{it} + \Delta \mu_{it} \tag{3}$$

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

$$\hat{MA}_{it} = \hat{\gamma_0} + \hat{\gamma}_X X_{it} \tag{4}$$

$$\hat{LA}_i = \frac{1}{2}\left[\left(MA_{i1} - \hat{MA}_{i1}\right) + \left(MA_{i2} - \hat{MA}_{i2}\right)\right] \tag{5}$$

With these estimated proxies of latent abilities, we estimate the effects of variables that remain constant through a child's life between 12 and 15 years old and which may determine latent ability using the YL sample. Since we are using two databases we require that these variables were available both in the YL questionnaire as well as for the ENHAB questionnaire. This allows 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).

$$\hat{LA}_i = \gamma_0^{LA} + \gamma_1^{LA} Z_i + \mu_i^{LA} \tag{6}$$

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.

$$\hat{\hat{LA}}_i = \hat{\gamma}_0^{LA} + \hat{\gamma}_1^{LA} Z_i \tag{7}$$

An assumption in this "matching" procedure is that the YL and ENHAB samples share similar characteristics such nationally representativeness.[4] With this prediction we estimate the wage equation and we analyze the gender wage gap as the theoretical model suggests: modelling wages as a function (basically) of latent cognitive and non cognitive abilities. The usual empirical approach was to model wages as a function of measured abilities (test scores)

[4] We should consider that the YL sample ignores children at the top 5% of the national income distribution.

which led to biased estimates. Thus, we will exploit the calculated proxies of latent abilities in ENHAB to compare the usual approach with this results.

Finally, we use the Oaxaca 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, lets consider the following relationship between wages and test scores:

$$LnW_i = \gamma_Y X_Y + \gamma_Y^C C + \gamma_Y^N N + \epsilon_Y$$

Estimating this equation provides biased estimators of γ^C and γ^N. In this equation latent abilities are unobserved and considered in the error term ϵ_Y. Since Cognitive (C) and non cognitive (N) test scores are functions of latent abilities, they are correlated with the error term. Therefore, estimated coefficients are not reproducing the effect of latent abilities on wages. Once the gender wage gap is identified and the proxies of latent cognitive and non cognitive abilities are estimated/predicted we will 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:

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

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

$$\bar{Y^1} - \bar{Y^0} = (\bar{X^1} - \bar{X^0})\hat{\beta^1} + \bar{X^0}(\hat{\beta^1} - \hat{\beta^0})$$

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 "two-fold decomposition". An extension of this method is the called "three-fold decomposition" which includes a third term that interacts (simultaneous) differences in observable characteristics and coefficients:

$$\bar{Y^1} - \bar{Y^0} = (\bar{X^1} - \bar{X^0})\hat{\beta^1} + \bar{X^0}(\hat{\beta^1} - \hat{\beta^0}) + (\bar{X^1} - \bar{X^0})(\hat{\beta^1} - \hat{\beta^0})$$

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

6 Results

In this section we compare the results of estimating the effect of cognitive and non cognitive abilities on wages by using measures of these skills (test scores) with those obtained by using proxies of latent ability. In each case we start by presenting the mincer equation of log wages controlling for schooling and abilities. Then, we apply the Oaxaca-Blinder Decomposition method in order to estimate the impact of abilities on the gender wage gap. Finally, in order to disentangle the effect of abilities on the gender wage gap 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. In order to proceed in this manner, we apply the procedure explained previously to obtain proxies of cognitive and non cognitive 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 III shows the results of a basic Mincer equation under the naïve assumption that there is no correlation between measured skills and schooling. Column 1 shows that after controlling for work experience, place of residence and mother tongue, an additional year of schooling leads to a 10.9% increase in log earnings. Column 2 controls for parent's schooling as this may explain part of the correlation between earnings and schooling. As suspected, the point estimate drops 15cognitive ability and emotional stability lead to higher wages while agreeableness and consistency of effort seem to reduce it. Including test scores also results in a reduction on the return to schooling which suggests that, in fact, the coefficient on column 2 overestimates the effect of schooling on earnings.

Table III: Mincer Equation with Measured Abilities

	[1]	[2]	[3]
Schooling	0.1091***	0.0934***	0.0804***
Experience	0.0255	0.0318*	0.0329*
Experience^2	−0.0003	−0.0004	−0.0004
Lives in Lima	0.0397	0.0297	0.0091
Mother Tongue (Spanish)	0.2675	0.4296	0.5023**
Father's Educational Level		0.0148	0.0097
Mother's Educational Level		0.0386**	0.0365**
Goldberg, Extraversion			0.0257
Goldberg, Agreeableness (kindness)			−0.0371
Goldberg, Agreeableness (cooperation)			−0.0616*
Goldberg, Conscientiousness (strong)			−0.005
Goldberg, Emotional Stability			0.0810***
Goldberg, Openness			−0.0069
Grit 2, Consistency of interest			−0.0453*
Grit 2, Persistence effort			0.0031
Cognitive Test Score			0.0751***
Constant	−0.6028**	−0.9125**	−0.8131***
Observations	1079	1079	1073
R-squared	0.157	0.17	0.187

NOTE: Robust standard errors in parentheses are clustered at a regional level. * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level. Sample is all people who are working at the moment of th interview and has available information on test scores and individual controls.

6.1.1 Oaxaca-Blinder Decomposition

In order to estimate the contribution of measured abilities on the gender wage gap we apply the Blinder-Oaxaca Decomposition. In our sample, the mean of log hourly wages is 1.417 for men and 1.141 for women, yielding a statistically significant wage gap of 0.276. Wage gap can be atributed 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 regresors) as men, more than 80% of the gender wage gap would be reduced if women shared the men's coefficients, given their own characteristics. The gender wage gap as well as the endowment effect and the differences in coeficients are significant even after controlling for standard individual characteristics. Columns 1 and 2 of Table IV 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.

A second line of analysis refers to the contribution of the individual predictors to the explained part of the gender gap. From now on, we will work with the composite measure of GRIT as the representative test for measuring non cognitive ability.[5] Table IV presents the portion of the wage gap in regresors and coefficients atributed to cognitive and non cognitive measured abilitiess. All four approximations show evidence that the "endowment effect" is

[5]I chose to work with GRIT as literature on non cognitive 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 been performed also with the measures of the rest of personality traits arriving to similar results.

due, basically, to different levels of (measured) cognitive abilities between men and women. No statistically significant effect were observed as a result of non cognitive (measures) abilities. In the case of differences in the coefficients, these were attributed to the individual controls or heterogeneity (between genders) in other unobserved characteristics.

Table IV: Oaxaca Decomposition with Measured Abilities

	[1]	[2]	[3]	[4]
Wage Gap	0.276***	0.276***	0.276***	0.276***
	(0.065)	(0.051)	(0.065)	(0.067)
Endowment	0.059***	0.045**	0.067***	0.058**
	(0.019)	(0.019)	(0.024)	(0.028)
Return	0.217***	0.231***	0.225	0.242***
	(0.063)	(0.048)	(0.067)	(0.062)
		Endowment		
Cognitive	0.059***	0.058***	0.067***	0.067***
	(0.019)	(0.015)	(0.023)	(0.023)
Non Cognitive	0.000	0.000	0.000	0.000
	(0.002)	(0.001)	(0.002)	(0.001)
		Return		
Cognitive	−0.171	−0.192	−0.163	−0.184
	(0.167)	(0.130)	(0.160)	(0.144)
Non cognitive	0.007	−0.017	0.007	−0.017
	(0.143)	(0.165)	(0.144)	(0.154)
Observations	1081	1081	1081	1081
Controls	No	Yes	No	Yes
Interactions	No	No	Yes	Yes

NOTE: Robust standard errors in parentheses are clustered at a regional level. * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level. Sample is all people who are working at the moment of th interview and has available information on test scores and individual controls. Controls include age (and its square) and residence in Lima.

The Oaxaca-Blinder decomposition could be approximated to a simple lifecycle model from a latent perspective. Cognitive and non cognitive abilities determine the schooling level. However, latent abilities are not observed. What is observed are the test scores as proxies of these abilities. Therefore, the returns estimated in the Oaxaca Blinder decompositions are functions of the return to schooling and the parameters governing the schooling level choice. Also, it is plausible that parameters influencing occupational choice are also affecting the overal estimated return.

6.1.2 Joint Estimation: Schooling, Employment, Occupation and Wages

The previous results provide information on the role of measures of cognitive and non cognitive skills on wages but don't account for other choices made by the individual before receiving a certain wage. In order to disentangle the effect of measured skills on each of this choices, we proceed with a joint estimation that considers choices of schooling, employment and occupation. The model follows an individual's line of choice: the individual is aware of its own level of abilities and uses this information to choose the years of schooling she is able to complete, then chooses to enter or not the labor market and thus, is employed or not. Once she decides to participate in the labor market she chooses the occupation and, finally, receives a certain wage according to her skills. For this analysis we work with two educational levels (complete secondary studies being the cutting point), being employed and two occupational choices ("white collar" or high skilled labor, and "blue collar" or low skilled labor).

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

$$l_i = \underbrace{L_{si}(\theta_S|LA_i)}_{\text{Schooling}} \overbrace{L_{hi}(\theta_h|LA_i, s_i)}^{\text{Working}} \underbrace{L_{oi}(\theta_O|LA_i, s_i, h_i = 1)}_{Occupation} \overbrace{L_{wi}(\theta_W|LA_i, s_i, h_i = 1, o_i)}^{Wages} \quad (8)$$

Each column of table V corresponds to each of the choices involved in the model. Results indicate that while measured cognitive skills seem to matter more in determining years of schooling and occupational choice, measured non cognitive abilities gain relevance for wages and employment. In terms of intergender differences, men seem to have higher returns to non cognitive 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, this estimated contributions consider measured abilities, which could be capturing the effect of other factors correlated with the outcome variables and measured abilities.

Table V: Joint Likelihood with Measured Abilities

Models	**Employment**	**Hourly Wages**	**Schooling**	**Occupational Choice**
Cognitive (Females)	−0.004*	-0.030	0.046***	0.043***
Interaction w/ Cognitive	0.004	0.040*	−0.006*	-0.007
Non Cognitive (Females)	0.122**	0.996**	0.190***	0.129
Interaction w/ Non Cognitive	0.195**	1.335**	0.134	0.105
Observations	2421			

NOTE: * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level. The interaction is (Male)*(Measured Ability).

6.2 Wages and Latent Abilities

In order to properly estimate the contribution of abilities on the gender wage gap, we consider latent abilities. In the following sections we present the results of the proposed procedure for estimating latent cognitive and non cognitive abilities and, then, reestimate the previous models with this resulting proxies of latent abilities.

6.2.1 Proxying Latent Abilities

Table VI shows the results of the first stage, where columns 1 to 3 correspond to cognitive ability (PPVT scores), self-esteem and self efficacy, respectively. Each regression controls for the child's caregiver's (corresponding) measured non cognitive ability[6], child's standardized height for age, standardized body mass index, age, perceived household status, household's wealth index, an indicator for having missed school for more than one week due to illness (-not so- exogenous variation in schooling) and log real consumption per capita. Standard errors are clustered by community.

In the case of self-esteem, changes in caregiver's self-esteem, height for age, body mass index, "having missed school for more than one week due to illness" and perceived household socioeconomic status are statistically significant, this latter having the larger effect. The results for self efficacy are similar but changes in perceived household socioeconomic status lacks significance. In terms of cognitive measured ability, variation in all the included controls except for the wealth index were statistically significant, with height for age, body mass index and "having missed school for more than one week due to illness" having a negative impact on test scores.

Table VI: First Stage Estimation (Fixed Effect Model of Measured Ability)

	Self Esteem	**Self Efficacy**	**Cognitive Ability**
Caregiver's Ability	0.039***	0.099***	-
	(0.013)	(0.023)	-
Wealth index	0.027	0.248	-2.793
	(0.180)	(0.261)	(1.854)
Standardized Height-for-age	−0.136***	0.187***	−1.698***
	(0.047)	(0.038)	(0.587)
Standardized Body-mass-index	0.068*	0.097	−1.377***
	(0.023)	(0.075)	(0.302)
Age in months	-0.002	0.001	0.793***
	(0.001)	(0.001)	(0.013)
(Perceived) Household Status	0.343***	0.007	1.144***
	(0.035)	(0.039)	(0.376)
Missed school due to illness	−0.140***	−0.271***	−1.416***
	(0.048)	(0.049)	(0.462)
Log Household Consumption	0.087	-0.043	-0.363
	(0.060)	(0.062)	(0.553)
Observations	349	349	349
R-squared	0.035	0.025	0.820

Note: Sample of children living in urban area with available information on relevant variables. * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level of the mean test between males and females. Clustered standard errors at community level.

Table VII shows the results of the second stage of our procedure. we control for gender and parents' educational level. Literature on skill formation suggest that latent ability is innate ability 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 ability formed up to when the child was 12 years old, so one could expect, a priori, that variables that are fixed until that moment should be important for determining this latent ability. Covariates such as parents educational level, gender and child's first language should be important, but

[6] No measure of the caregiver's cognitive ability was available in the dataset.

not others such as characteristics of secondary education (which would also be endogenous). This is what motivates the reduced model. For the proxies of all three latent ability, all included controls[7] are statistically significant. While women appear to have a higher endowment of self efficacy, the opposite is the case for cognitive abilities. Almost consistently, parent's education has a positive impact on all three proxies.

Table VII: Second Stage Estimation (Latent Ability on Permanent Characteristics, YL)

	Self Esteem	**Self Efficacy**	**Cognitive Ability**
Male	0.012	−0.361***	1.026***
	(0.016)	(0.013)	(0.361)
Father's schooling	0.035***	0.018***	1.442***
	(0.006)	(0.004)	(0.075)
Mother's schooling	0.016***	−0.008**	0.461***
	(0.002)	(0.003)	(0.090)
Constant	−0.581***	0.43***	−61.802***
	(0.071)	(0.044)	(1.326)
Observations	349	349	349
R-squared	0.048	0.067	0.169

NOTE: * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level of the mean test between males and females. Clustered standard errors at community level.

Table VIII: Third Stage Statistics (Predicted Latent Abilities in both databases)

Young Lives Sample	**Female**	**Male**	**Gender Gap**	**Whole Sample**
Latent Cognitive Ability	-42.676	-40.854	1.821	-41.716
			(1.462)	
Latent Self Efficacy	0.532	0.182	-0.350***	0.347
			(0.073)	
Latent Self Esteem	-0.067	-0.036	0.031	-0.050
			(0.074)	
Observations	165	184		349
ENHAB Sample	**Female**	**Male**	**Gender Gap**	**Whole Sample**
Latent Cognitive Ability	-47.409	-46.209	1.201***	-46.829
			(0.201)	
Latent Self Efficacy	0.516	0.156	−0.361***	0.342
			(0.001)	
Latent Self Esteem	-0.198	-0.193	0.005	-0.196
			(0.005)	
Observations	3872	3627		7499

NOTE: Predicted fixed effects for the ENHAB sample were based on YL estimates for the urban subsample. Adjusted standard errors for intra-group correlation are reported in parentheses; * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level.

Since both surveys are nationally representative, the matching procedure applied should be plausible. Table VIII shows descriptive statistics of both fixed effect predictions (for the YL and ENHAB samples) in the full sample and by gender. As can be seen, both predictions share similar characteristics, thus, our procedure seems valid.

[7]Mother tongue was omitted because more than 90% of the sample has spanish as their mother tongue, thus, it lacks variation.

Finally, we estimate the effect of latent abilities on wages. Table IX compares the results of the basic Mincer equation obtained by including measured abilities (column 3) and those obtained by controlling, instead, for our predicted latent abilities[8] (column 4). Two results are worth highlighting. First, the return to schooling in column 4 is much larger than that of column 3 and more similar to that of column 2. 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 non cognitive abilities. This evidences that there is an effect of non cognitive abilities on wages, but also that this now captures also the indirect effect of abilities through schooling now that we are able to control for both schooling and latent abilities.

Table IX: Mincer Equation with Latent Abilities

	[1]	[2]	[3]	[4]
Schooling	0.1091***	0.0934***	0.0804***	0.0914***
Experience	0.0255	0.0318*	0.0329*	0.0333**
Experience^2	−0.0003	−0.0004	−0.0004	−0.0004
Lives in Lima	0.0397	0.0297	0.0091	0.0606*
Mother tongue (Spanish)	0.2675	0.4296	0.5023**	0.2975
Father's schooling		0.0148	0.0097	0.0507
Mother's schooling		0.0386**	0.0365**	0.0312
Goldberg, Extraversion			0.0257	
Goldberg, Agreeableness (kindness)			−0.0371	
Goldberg, Agreeableness (cooperation)			−0.0616*	
Goldberg, Conscientiousness (strong)			−0.005	
Goldberg, Emotional Stability			0.0810***	
Goldberg, Openness			−0.0069	
Grit 2, Consistency of interest			−0.0453*	
Grit 2, Persistence effort			0.0031	
Cognitive Test Score			0.0751***	
Predicted (Latent) Non Cognitive Ability				−0.7259***
Predicted (Latent) Cognitive Ability				−0.0044
Constant	−0.6028**	−0.9125**	−0.8131***	−0.8876
Observations	1079	1079	1073	1079
R-squared	0.157	0.17	0.187	0.193

NOTE: Robust standard errors in parentheses are clustered at a regional level. * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level. Sample is all people who are working at the moment of th interview and have available information on test scores and individual controls.

6.2.2 Oaxaca-Blinder Decomposition

This section describes the results obtained after applying the Oaxaca-Blinder Decomposition to the whole sample of working age population[9] but accounting for differences in latent ability (the previously estimated proxies). As stated in the specification with measured abilities, a

[8] From now on we will use self efficacy as the representative non cognitive ability. This may be debatable, but for the case of self-esteem descriptive statistics on the YL sample and estimated as well as predicted latent abilities showed no relevant gender wage gaps.

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

significant gender wage gap exists. In contrast to the results obtained in the previous section, the gap found with this specification is not only attributed to differences in returns or endowments in cognitive skills, but in this case non cognitive skills also play an important role in explaining the gender wage gap.

Table X shows the results corresponding to the Oaxaca-Blinder Decomposition for the ENHAB sample accounting for differences in latent cognitive ability as well as latent self efficacy (as proxy of non cognitive latent ability). After applying the two-fold and three-fold approximations, data support the fact that the gender wage gap is atributed to group differences in the coefficients and the predictors (after adding standard controls, columns 2 and 4). Regarding differences in returns, we can see that men have a higher return on cognitive skills increasing the gender wage gap but no significant differences in returns to non cognitive skills affect the gender wage gap. In terms of differences in endowment of abilities, table X shows that differences in cognitive and non cognitive ability favor men regarding earnings. While the higher endowment of cognitive ability amongst men contribute to the gender wage gap, it seems to appear that a if women had the same endowment of non cognitive ability as men, they would earn significantly higher wages. At first sight, this seems to contradict what was shown in the initial descriptive statistics where women had higher levels of self efficacy. Nonetheless, this is explained by the fact that the return to non cognitive ability for women is negative (as we would have noticed if we had estimated the Mincer equation with latent abilities by gender). Thus, a higher endowment of a skill that has a negative return in fact impact negatively on women incresing the gender wage gap.

Table X: Oaxaca-Blinder Decomposition with Latent Abilities

	[1]	**[2]**	**[3]**	**[4]**
Wage Gap	0.241***	0.241***	0.241***	0.241***
	(0.032)	(0.027)	(0.032)	(0.035)
Endowment	0.419***	0.607***	0.458***	0.618***
	(0.117)	(0.134)	(0.148)	(0.152)
Return	−0.178	−0.366***	−0.160	−0.367***
	(0.124)	(0.136)	(0.157)	(0.141)
	Endowment			
Cognitive	0.026***	0.037***	0.034***	0.043***
	(0.008)	(0.010)	(0.011)	(0.013)
Non Cognitive	0.393***	0.580***	0.424***	0.585***
	(0.115)	(0.132)	(0.144)	(0.148)
	Return			
Cognitive	0.540***	0.477*	0.544***	0.481**
	(0.178)	(0.254)	(0.180)	(0.187)
Non cognitive	0.049	0.004	0.063	−0.001
	(0.186)	(0.272)	(0.265)	(0.270)
Observations	4079	4079	4079	4079
Controls	No	Yes	No	Yes
Interactions	No	No	Yes	Yes

Note: Robust standard errors in parentheses are clustered at the inividual level. * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level. Sample is all people who are working at the moment of the interview and has available information on individual controls. Controls include age (and its square) and residence in Lima.

6.2.3 Joint Estimation: Schooling, Employment, Occupation and Wages

Table XI shows the results of the joint estimation considering cognitive and non cognitive latent abilities. Interpreting the role of both abilities in each of the choices considered lead to interesting results. First, cognitive ability is crucial for attaining higher levels of education and this is so for men and women in the same magnitude. Non congitive abilities don't seem determinant for this choice. The other three choices must be interpreted together. In general we observe that there are statistically significant intergender differences in the returns of cognitive and non cognitive abilities in all three outcomes. This is mainly driven by occupational choice and the negative return to non cognitive abilities combined with the fact that women have higher levels of this ability. We could interpret that men earn higher wages because when employed their equilibrium assignation is towards occupations with higher rewards toward cognitive skill. This, combined with the fact that men have higher cognitive skills, helps explain the gender wage gap. Interestingly enough, it seems to appear that the return to non cognitive ability in terms of employment and wages for non cognitive abilities is higher for men even though they present a lower endowment of this skill, which seems to be reasonable in terms of valuing what is uncommon.

Table XI: Joint Likelihood with Latent Abilities

Models	**Employment**	**Hourly Wages**	**Schooling**	**Occupational Choice**
Cognitive (Females)	-0.014^{***}	-0.076^{***}	0.055^{***}	0.065^{***}
	(0.002)	(0.018)	(0.004)	(0.004)
Interaction w/ Cognitive	-0.009^{***}	-0.055^{***}	-0.005	0.014^{***}
	(0.002)	(0.017)	(0.003)	(0.004)
Non Cognitive (Females)	-0.117	-3.013	0.109	-1.539^{***}
	(0.298)	(2.314)	(0.443)	(0.483)
Interaction w/ Non Cognitive	0.817^{**}	5.794^{**}	-0.515	-1.347^{***}
	(0.319)	(2.463)	(0.438)	(0.512)
Observations	**7499**			

NOTE: * indicates 10% significance level; ** indicates 5% significance level; and *** indicates 1% significance level. The interaction is (Male)*(Latent Ability).

In contrast with the results from the joint estimation using measured abilities, we can observe that intergender differences in cognitive and non cognitive abilities in favor of men drop when considering latent abilities. In the case of cognitive abilities, this gaps even turn significant in favor of women. Moreover, returns to both latent abilities gain significance for the occupational choice outcome. This supports our prior that most differences atributable to abilities occur within occupational choice.

6.3 Robustness Check: Common Support

In order to obtain proxies for latent cognitive and non cognitive abilities, the proposed procedure required an imputation method from an estimation in another database. As much as descriptive statistics were presented regardin the estimated and predicted latent abilities for the YL and ENHAB samples, respectively, empirical literature suggests that one should work with the common support. The common support is determined by the common area of both distributions of estimated/predicted latent abilities.

Figures 1 and 4 in the Appendix show the distributions of cognitive and non cognitive latent abilities, respectively, for the YL and ENHAB samples for women and men. In every case, the YL sample distributions have a higher variance than the ENHAB sample, so the common support bounds only the YL sample. Being such the case, we calculate the higher and lower values of each of the distribution for the ENHAB sample and reestimate the fixed effects with the YL subsample whose initial fixed effect was comprised in the common support. Then, we predict again to the ENHAB sample and compare the distribution of fixed effects for the common support. As figures 7 and 10 (in the Appendix) evidence, there is no important gain of proceding in such a manner. Moreover, this bounding of the YL sample implies dropping a significant amount of observations and thus impacting negatively in the precision of the estimates obtained by means of the common support.

Having observed that slimming the sample to the common support doesn't contribute to better estimations, but does completely the opposite we maintain the analysis so far estimated by means of the whole sample of YL children with available information on the key variables.

7 Conclusions

This paper presents preliminary evidence on the role of cognitive and non cognitive skills in closing the gender wage gap. In a first attempt to estimate their effect on wages, we followed the basic empirical approach of modelling wages in terms of measured ability. In addition, we complement my work by applying a procedure that allows to estimate proxies of latent abilities, thus being able to estimate a model in terms of latent ability. we base 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 non cognitive skills. In addition, we estimate a joint model of schooling, employment, occupational choice and earnings in order to disentangle the effects of latent abilities in the gender wage gap throughout an individual's choices previous to earning a certain wage. My main contribution is analyzing the role of latent non cognitive and cognitive abilities to the gender wage gap in a developing country by estimating and accounting for proxies of latent abilities and disentangling the effect of these abilities by means of a joint model of schooling, employment, occupational choice and earnings.

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 evidences significant intergender differences in the endowment of cognitive skills but no relevant differences in terms of non cognitive abilities (in endowment or returns). Estimating the joint model evidences that differences in non cognitive abilities between men and women are important but on choices prior to wage determination. Cognitive skills seem to be relevant in determining years of schooling and occupational choice and measured non cognitive ability for wages and employment.

The application of the fixed effect model to identify proxies of latent cognitive and non cognitive abilities leads to a clearer relationship between abilities and wages. As a result of the procedure of estimation of latent abilities we am able to obtain a more precise application of the gender wage gap decomposition. When estimating a wage equation in terms of measured ability (due to the unavailability of actual values for latent ability), one would obtain biased estimates due to the dependance of these on schooling and the endogenous nature of this latter variable. After being able to identify proxies for these latent abilities, one is left with ubiased estimates and thus, the real contribution of skills to wages (so far, conditional on occupational choice). Latent abilities revealed to be highly statistically significant for mean wages as well as accounting for inter-gender differences. In particular differences in the endowment of non cognitive abilities contribute (negatively) to the gender wage gap. Women appear to be receiving lower wages than men because they have higher endowments of non cognitive skills which are negatively valued (punished) in the labor market. No significant intergender differences in the returns to cognitive abilities were found. Moreover, the estimation of the joint model sheds light towards the fact that the observed gender wage gap is mainly attributed to differences within occupational choice. Cognitive and non cognitive abilities are valued differently for men and women in terms of schooling, employment and wages, but basically men seem to earn higher wages because their equilibrium assignation is towards occupations with higher rewards to cognitive skills, which they are most endowed with.

In sum, we am able to conclude that the proposed procedure for estimating latent abilities , not free from numerous limitations, leads to reasonable results in light of related empirical literature on the relationship between wages, schooling and abilities.

Further extensions to this paper include explore deeper into the marginal effects of latent abilities on schooling, employment, occupation effects and wages, explore gender wage gap by income quantiles and compare my results with Heckman, Stixrud and Urzua's (2006) original identification strategy based in different test scores for the same skill in a same time period.

References

Bassi, Marisa and Sebastian Galiani, "Labor Market Insertion of Young Adults in Chile," 2009. Inter Amercian Development Bank, unpublished manuscript.

Cobb-Clark, Deborah A. and Michelle Tan, "Noncognitive Skills, Occupational Attainment, and Relative Wages," IZA Discussion Papers 4289, Institute for the Study of Labor

(IZA) July 2009.

Cunha, Flavio and James Heckman, "TThe Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

_ **and James J. Heckman**, "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 2008, *43* (4).

_ **,** _ **, and Susanne M. Schennach**, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 05 2010, *78* (3), 883–931.

Díaz, Juan Jose, Omar Arias, and David Vera Tudela, "Does Perseverance Pay as Much as Being Smart?: The Returns to Cognitive and Non-cognitive Skills in urban Peru," 2012. Mimeo.

Fortin, Nicole M., "Greed, Altruism and the Gender Wage Gap," Technical Report, University of British Columbia February 2006.

_ , "The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People," *Journal of Human Resources*, 2008, *43* (4).

Grove, Wayne A., Andrew Hussey, and Michael Jetter, "The Gender Pay Gap Beyond Human Capital: Heterogeneity in Noncognitive Skills and in Labor Market Tastes," *Journal of Human Resources*, 2011, *46* (4), 827–874.

Hansen, Karsten T., James J. Heckman, and K.J.Kathleen J. Mullen, "The effect of schooling and ability on achievement test scores," *Journal of Econometrics*, 2004, *121* (1-2), 39–98.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, July 2006, *24* (3), 411–482.

Helmers, Christian and Manasa Patnam, "The formation and evolution of childhood skill acquisition: Evidence from India," *Journal of Development Economics*, 07 2011, *95* (2), 252–266.

Jacob, Brian A., "Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education," NBER Working Papers 8964, National Bureau of Economic Research, Inc 05 2002.

Urzúa, Sergio, "Racial Labor Market Gaps: The Role of Abilities and Schooling Choices," *The Journal of Human Resources*, 2008, *43* (4), 919–971.

_ **, David Bravo, and Claudia Sanhueza**, "Ability, Schooling Choices and Gender Labor Market Discrimination: Evidence for Chile," in Hugo Ñopo, Alberto Chong, and Andrea Moro, eds., *Discrimination in Latin America: An Economic Perspective*, 2009.

8 Appendix

8.1 Psychosocial Components - Young Lives Sample

Table XII: Round 2

Variables	Sample Mean Mean	SD	Males Mean	SD	Females Mean	SD
Child's self efficacy						
If we try hard we can improve my situation in life	1.987	0.157	1.981	0.188	1.994	0.113
Other people in my family make all the decisions about how we spend my time	1.267	0.941	1.320	0.924	1.206	0.959
I like to make plans for my future studies and work	1.943	0.327	1.934	0.349	1.953	0.300
If we study hard we will be rewarded with a better job in the future	1.957	0.277	1.956	0.275	1.959	0.280
I have no choice about the work we do	0.859	0.976	0.841	0.978	0.887	0.979
Child's self-esteem						
I feel proud to show my friends where we live	1.800	0.590	1.800	0.585	1.801	0.597
I am ashamed of my clothes	0.257	0.664	0.232	0.635	0.286	0.696
I am ashamed of my shoes	0.188	0.581	0.157	0.537	0.223	0.628
I feel proud of the job done by the head of household	1.878	0.468	1.896	0.439	1.857	0.500
I am often embarrassed because we do not have the right supplies for school	0.453	0.829	0.439	0.817	0.468	0.843
I am worried that we don't have the correct uniform	0.698	0.951	0.652	0.936	0.752	0.967
I am proud of my achievements at school	1.928	0.356	1.909	0.401	1.949	0.295
I am embarrassed by the work we have to do	0.196	0.587	0.150	0.521	0.268	0.675
The job we do makes me feel proud	1.826	0.555	1.885	0.458	1.732	0.675
Caregiver's self efficacy						
If we try hard we can improve my situation in life	1.942	0.315	1.962	0.242	1.918	0.382
I like to make plans for my future	1.902	0.416	1.902	0.406	1.902	0.429
I have no choice about which school to send my child to	0.836	0.984	0.777	0.973	0.905	0.994
If my child gets sick we can do little to help him/her get better	0.456	0.836	0.405	0.803	0.516	0.871
I can do little to help my child do well in school no matter how hard we try	0.549	0.887	0.472	0.842	0.637	0.930
Caregiver's self-esteem						
I feel proud to show my friends or other visitors where we live	1.782	0.601	1.783	0.592	1.782	0.612
I am ashamed of my clothes	0.344	0.746	0.302	0.707	0.392	0.787
I feel proud of the job done by the hh head	1.896	0.429	1.879	0.456	1.917	0.395
The job we do makes me feel proud	1.939	0.328	1.943	0.313	1.934	0.346
I feel proud of my children	1.972	0.231	1.970	0.237	1.975	0.224

Table XIII: Round 3

Variables	Sample Mean Mean	Sample Mean SD	Males Mean	Males SD	Females Mean	Females SD
Child's self efficacy						
If we try hard we can improve my situation in life	1.191	0.544	1.155	0.554	1.232	0.531
Other people in my family make all the decisions about how we spend my time	3.340	0.904	3.383	0.914	3.290	0.892
I like to make plans for my future studies and work	1.137	0.660	1.039	0.684	1.249	0.613
If we study hard we will be rewarded with a better job in the future	1.359	0.588	1.330	0.608	1.392	0.564
I have no choice about the work we do	0.224	0.992	0.111	1.012	0.344	0.957
Child's self-esteem						
I am proud of my clothes	0.944	0.655	0.966	0.632	0.919	0.680
I am proud of my shoes/chappals or of having shoes/chappals	0.917	0.759	0.921	0.755	0.913	0.765
I am never embarassed because we do not have the right books, pencils or other eq	0.618	0.940	0.602	0.943	0.635	0.938
I am proud that we have the correct uniform	1.099	0.605	1.060	0.584	1.144	0.626
I am proud of the work I have to do	1.042	0.572	1.052	0.571	1.031	0.574
Caregiver's self efficacy						
If we try hard we can improve my situation in life	1.042	0.461	1.048	0.483	1.036	0.436
I like to make plans for my future	0.578	0.837	0.545	0.803	0.616	0.873
I have no choice about which school to send my child to	0.578	0.837	0.545	0.803	0.616	0.873
If my child gets sick we can do little to help him/her get better	0.869	0.689	0.915	0.676	0.817	0.702
I can do little to help my child do well in school no matter how hard we try	0.767	0.811	0.847	0.800	0.675	0.816
Caregiver's self-esteem						
I feel proud to show my friends or other visitors where we live	0.880	0.643	0.904	0.645	0.852	0.641
I am ashamed of my clothes	0.736	0.660	0.734	0.705	0.740	0.606
I feel proud of the job done by the hh head	0.965	0.688	0.943	0.726	0.990	0.644
The job I do makes me feel proud	1.030	0.556	0.997	0.618	1.068	0.474
I feel proud of my children	1.439	0.581	1.398	0.575	1.486	0.584

8.2 Latent Ability Distributions and Common Support

Figure 1: Cognitive Ability Distribution

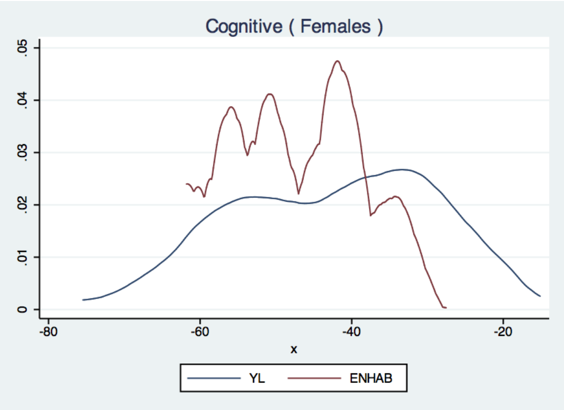

Figure 2: Females

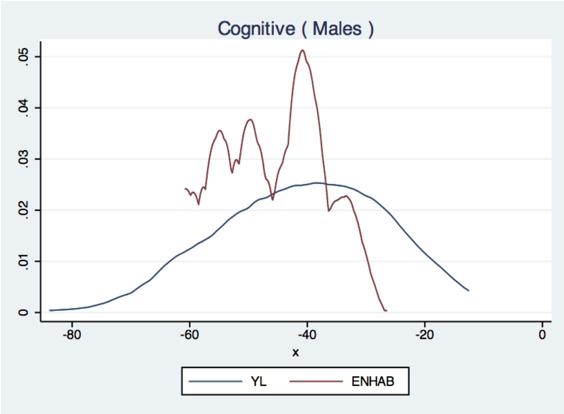

Figure 3: Males

Figure 4: Non cognitive Ability Distribution

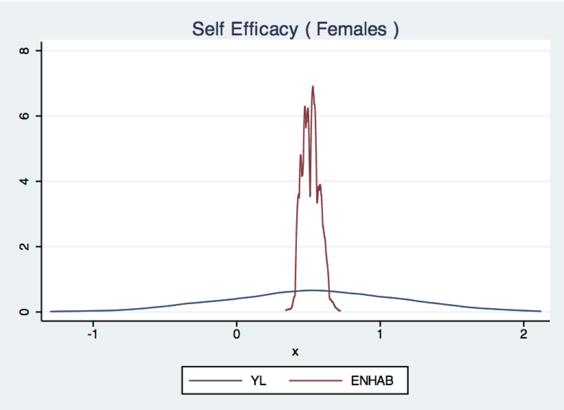

Figure 5: Females

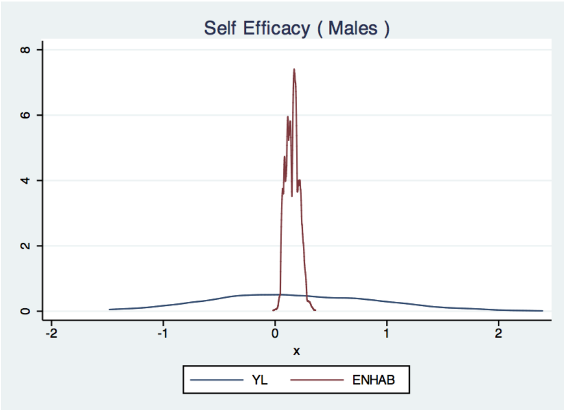

Figure 6: Males

Figure 7: Cognitive Ability Distribution (Common Support)

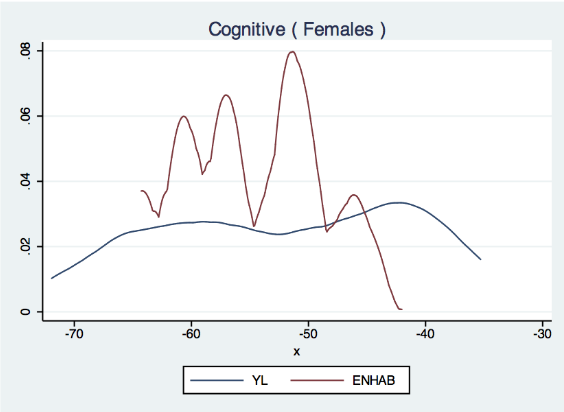

Figure 8: Females

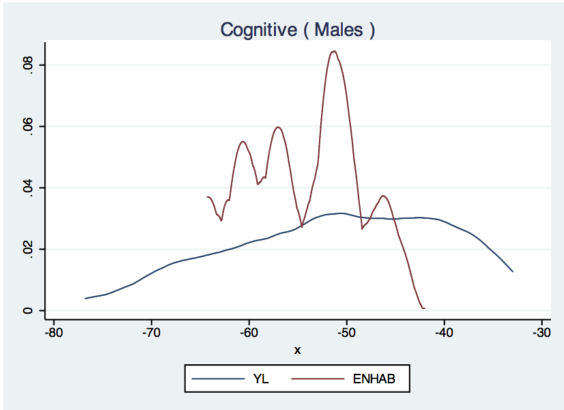

Figure 9: Males

Figure 10: Non cognitive Ability Distribution (Common Support)

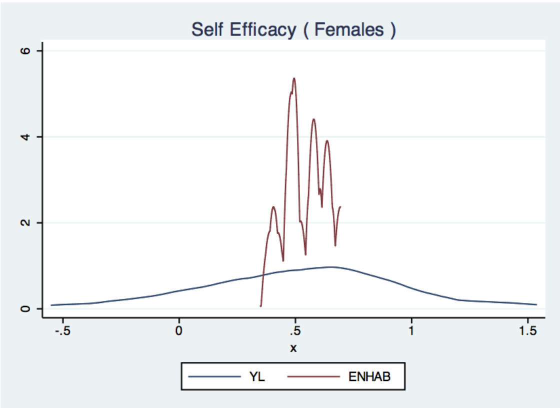

Figure 11: Females

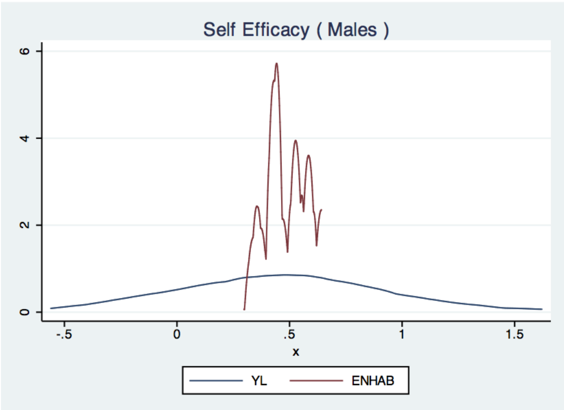

Figure 12: Males

Cognitive and Non-cognitive Skills and Wages: The Role of Latent Abilities on the Gender Wage Gap in Peru

Recently there has been growing interest in the relationship between cognitive and non-cognitive abilities and labour market outcomes. A large literature provides evidence on the positive connection between cognitive test scores and higher wages. Fewer and newer papers have explored the correlation between non cognitive test scores and wages. However, attention is focused on developed countries. Test scores suffer two limitations. First, they can be considered outcomes of the schooling level and latent (unobserved) cognitive and non-cognitive abilities. Second, they are potentially measured with error. The main objective of this paper 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 Young Lives Study (YL) and the Peruvian Skills and Labor Market Survey (ENHAB). Young Lives has panel data information on test scores and ENHAB has information on test scores and wages. Results show that even though when accounting for measured abilities differences in non-cognitive abilities seem irrelevant, when accounting for differences in actual latent ability non cognitive 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 not only in wage profiles, but in schooling, employment and occupation decisions.

Young Lives
An International Study of Childhood Poverty

Contact:
Young Lives
Oxford Department of International Development,
University of Oxford,
3 Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@younglives.org.uk
Website: www.younglives.org.uk