Does Perseverance Pay as Much as Being Smart?: The Returns to Cognitive and Non-cognitive Skills in urban Peru[1]

Juan José Díaz[2]
GRADE

Omar Arias[3]
World Bank

David Vera Tudela[4]
World Bank

November 30, 2013

Abstract

This paper estimates the returns to cognitive and socio-emotional ("non-cognitive") skills using a labor force survey in Peru designed to measure these skills in the working-age (14-50) urban population, the first of its kind in a developing country. The survey measures a wide range of cognitive (Peabody receptive language, verbal fluency, working memory, and numeracy/problem-solving) and personality traits to proxy for socio-emotional skills (Big-Five Factors, Grit). We corroborate findings from developed countries that both types of skills are significant correlates of earnings. Using data on instrumental variables to address the potential endogeneity of measured skills vis-à-vis schooling, the findings confirm that socio-emotional and cognitive skills are equally valued in the Peruvian labor market. A one standard deviation change in an overall cognitive skill measure and in the perseverance facet of Grit each generates a 9% increase on average earnings, conditional on schooling. The effect size of an increase in years of schooling (of about 3 years) is a 15% increase in earnings, conditional on skills. The returns to other socio-emotional skills vary across dimensions of personality: 5% higher earnings for emotional stability while 8% lower earnings for agreeableness.

JEL CODES: J24, I24
Key words: returns to cognitive, non-cognitive skills, schooling.

---

[1] The authors would like to acknowledge Santiago Cueto, Mary Claux and Maria Isabel de la Rosa who led the development of the instruments and measures of skills used in the paper, as part of the Skills and Labor Market Survey developed by the World Bank in Peru in 2011, led by one of the authors. We are thankful to Angela Duckworth for kindly facilitating the use of the Grit Scale as well as to Nancy Guerra and Brent Roberts for invaluable suggestions during the design of the Big-Five personality inventory. We benefitted from insights and comments from Flavio Cunha, Wendy Cunningham, Rodrigo Soares and participants in seminars at the World Bank, the Inter-American Development Bank and the University of Maryland in presentations of earlier versions of the paper. The opinions expressed herein do not represent the views of the organizations to which the authors are affiliated. All remaining errors are the sole authors' responsibility.
[2] Principal Researcher, *Grupo de Análisis para el Desarrollo* - GRADE.
[3] Corresponding author: Lead Economist, Eastern Europe and Central Asia Human Development Department at the World Bank; email address: oarias@worldbank.org.
[4] Research Analyst, Country Department for Bolivia, Chile, Ecuador, Peru and Venezuela, the World Bank.

1. Introduction

An increasing number of empirical studies from psychology and economics consistently show that both cognitive and non-cognitive skills are important determinants of socioeconomic success. Cognitive skills are associated to the capacity to learn and solve problems. Non-cognitive skills are personality traits, such as perseverance, motivation, sociability and emotional stability.

Much of the recent empirical literature highlights the fact that non-cognitive skills are highly valued in the labor market. For instance, evidence from the Perry Preschool program, an early childhood intervention targeted to disadvantaged children with low IQ, show that the intervention improved non-cognitive abilities of program participants but did not affect their cognitive ability, and generated large impacts on schooling, labor market outcomes and other outcomes such as crime rates. This is particularly relevant for the design of policies aimed at fostering human capital because much of the efforts regarding education and job training to date concentrate on the production of cognitive skills.

The evidence also suggests that skills are formed over the lifecycle, and are affected by genetic endowments and the environment. Thus, families play an important role in the acquisition and production of skills. In addition, the formation of skills is characterized by sensible and critical periods and by complementarity. Investments to promote skills shall be made before reaching critical periods in order to increase the formation of later skills. Cognitive skills, achieve stability earlier at the lifecycle, and are fairly well developed between age 8 and 10, while non-cognitive skills remain malleable over a longer span. This is important for policy design as interventions to promote skill formation shall be implemented when they are needed and because remedial policies are usually ineffective and extremely costly.

In this paper we assess the relationship between labor earnings, schooling and cognitive and non-cognitive skills in urban labor markets from Peru. Our objective is to find what skills are valued in the labor market in order to inform the design of policies aimed at improving the employment and earnings capacity of the Peruvian labor force. As far as we know, this is one of only a few existing studies outside the developed world that assess the relationship between cognitive and non-cognitive skills and labor earnings at a national workforce scale.

We use data from a random sample of the working-age (14-50) urban population in Peru collected trough a survey specially designed to explore the relationship between labor outcomes, education and cognitive and non-cognitive skills. The survey instrument includes a battery of tests specially designed to measure cognitive skills (the Peabody Picture Vocabulary Test, and tests of verbal fluency, working memory, and numeracy/problem-solving) and non-cognitive skills or personality traits (the big-five factors of personality and the Grit personality traits of perseverance and the will to strive for long term goals). It also includes a module on schooling trajectories from pre-school to college, which provides information on parental background, family structure at young age, socioeconomic status while pursuing education, self-reported scholastic achievement and effort exerted at school, and parent's valuation of education. This allows us employing usually hard to find possible instrumental variables for schooling and measured abilities in a single dataset.

The paper proceeds as follows. Section 2 provides a brief summary of the evidence on the different dimensions of skills, the way these skills are formed, and the relationship to labor outcomes. Section 3 presents the econometric models used in our empirical analysis. Section 4 describes the data and the measured skills considered in the analysis. Section 5 presents an explorative analysis on the relationship between cognitive and non-cognitive skills, treated separately, and labor earnings. Section 6 reports the results of our analysis when we put all the skills indicators and schooling together. Section 7 concludes.

2. Literature

People embodied two broad types of skills, cognitive skills related to reasoning, planning, abstract thought and problem solving; and non-cognitive skills or personality traits, such as extraversion and emotional stability, related to perseverance, motivation, and self-control. All of these skills affect the way individuals make decisions and how do they perform in their lives.

Measurements of cognitive skills are rooted in theories of intelligence. Following the definition of intelligence provided by the American Psychology Association, cognitive skills are largely associated to "all forms of knowing and awareness such as perceiving, concerning, remembering, reasoning, judging, imagining, and problem solving." There is a consensus in the Psychology literature on the existence of two broad dimensions of intelligence. *Fluid intelligence* is the domain of raw problem-solving ability, while *crystallized intelligence* is the domain of knowledge to solve problems. Fluid intelligence is related to IQ, and the evidence suggests that its formation finish by age 8 or 10. Deficiencies generated during the process are carried along the individual's life. Crystallized intelligence, on the contrary, remains malleable over a longer life span beyond the age 10 and is affected by formal schooling as well as for other stimuli that enhance mental capacity.

Personality encompasses many traits such as perseverance, industriousness, reliability, cooperation, self-control, self-efficacy, self-esteem, security. Psychologists have produced a large array of batteries aimed at measuring such traits. In particular, economists have used the Rotter Locus of Control scale and the Rosenberg Self-Esteem scale to investigate the relationship between non-cognitive skills and schooling choice and employment.

More recently psychologists have developed a widely accepted taxonomy of personality known as the Five Factor Model or Big Five Factors of personality (Goldberg 1990) that identifies five broad dimensions of personality. These domains are openness, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN). Openness is associated to an appreciation for art, emotion, adventure, unusual ideas, curiosity, and variety of experience. Conscientiousness is a tendency to show self-discipline, act dutifully, and aim for achievement planned rather than spontaneous behavior. Extraversion is related to energy, positive emotions, and the tendency to seek stimulation in the company of others. Agreeableness is related to the tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. Neuroticism or emotional stability is related to the tendency to experience unpleasant emotions easily, such as anger, anxiety, depression, or vulnerability.

How skills are formed

During the mid fifties, economists began to study the relationship between investments on education and on-the-job training and earnings. The theory of human capital emerged as an explanation of these investments and the earnings distribution among individuals and over time. In this theory, ability is a single dimensional innate characteristic of the individual. Ability can influence school choice as more able people face lower costs to acquire education. In addition, ability influence earnings directly as more able people are thought to be more productive, and indirectly as more able people have more education which also increases productivity. Human capital on the contrary can be produced and increased over the life cycle by education and trainings investments (Ben Porath, 1964; Mincer, 1958).

In recent years, based on evidence provided by psychology and economics, economists have started to expand models of human capital by recognizing that skills and ability are multi-dimensional characteristics of the individual and that these skills are not immutable but can be affected by families, schools, and firms (Cunha 2005, Cunha and Heckman 2006. Even more, collaboration between psychologists and economists is beginning to enrich the way we interpret the evidence on cognitive and non-cognitive skills, as well as how can we model and explain the process of skill formation (Duckworth et al 2008).

A first crucial issue to put forward is that skills are produced over the life cycle of the individual by their families, schools, environments and workplace. In general most attention is placed on formal schooling by policymakers, but much of the non-cognitive skills and motivation are shaped by other informal or non institutional sources of learning.

In addition, a second crucial issue is that skills differ in their malleability over the life cycle. The available evidence from cognition studies suggest that cognitive ability is formed relatively early in life and becomes less malleable as children age. In particular, it seems that by age 8 or 10, intelligence or cognitive ability is fairly well set. On the contrary, non-cognitive ability remains malleable for a longer span over the life cycle.

There are different stages in the life cycle that are critical to the formation of some kinds of skills, as emphasized in Shonkoff and Phillips (2000). In addition, the production of skills at earlier stages increases the productivity of skills formation raising the production of skills at later stages, which is termed *dynamic complementarities* in skill formation (Heckman 1999, Heckman and Carneiro 2003; Cunha, Heckman, Lochner and Masterov 2005; Cunha and Heckman 2006).

When children are exposed to deprived early environments which results in low skills, it is still possible to compensate latter on to increase skills provided there is enough substitutability between investments through time. If this is the case, there are two channels for compensation. The first is through increasing late skill investment, which raise the amount of skills in later periods. The second is through the choice of activities or tasks in the marketplace that are intensive in the skills abundant (or less scarce) for the individual. On the contrary, if investments are complementary and there are critical stages for the particular skills, late investments are ineffective and likely inefficient.

An additional issue is that when there is complementarity of investments on skills over time, later investments are required to match early investments in order to harvest the proceed of early investments. Under complementarity, if late investments fall short of early investments, then the production of skills in latter periods is reduced by the self-productivity effect.

Relationship to outcomes

Cognitive and non-cognitive skills are determinants of school choice, labor market performance as well as participation in risky behaviors such as crime, teenage pregnancy, drug use and other deviant activities (Heckman, Stixrud and Urzua 2006; Borghans et al 2008). Their predictive power on a variety of adult outcomes has been well established in the psychology literature (e.g., Roberts et al 2011; Roberts et al 2011). Cunha, Heckman, Lochner and Masterov (2005) survey evidence showing that cognitive ability affects the likelihood of acquiring higher levels of education and advanced training as well as the economic returns to these activities.

Earlier work by Bowles and Gintis (1976) and Edwards (1976) showed that some non-cognitive skills such as dependability and persistence are highly valued by employers. This evidence is confirmed by Klein, Spady and Weiss (1991), and more recently by the empirical work reported by Heckman, Stixrud and Urzua (2006) that addressed several limitations from previous literature such as reverse causality and measurement error.

Recent evidence also confirms that skills are highly valued by employers (Heineck and Anger 2008), and that employers assess cognitive and non-cognitive skills for hiring, promotion and wage setting policies (Farkas 1996; Jenkins 2001; Psacharopoulos and Schlotter 2009).

All of these findings come mainly from the U.S. and other developed countries, either from national surveys, employment records, or program pilot demonstrations. There is scarce evidence for Latin America's countries. Two recent studies from Chile are the exception. Urzua (2009) explores the relationship between skills and the transition from school to work. Bassi and Galiani (2009), also exploring young adult data from Chile, is the only study we have found for Latin America that addresses the relationship between skills and labor earnings.

Bassi and Galiani (2009) use a survey with nation-wide data for young adults between age 25 and 30. The survey instruments include two tests of cognitive skills and two tests of non-cognitive skills. They measure intellectual ability using the Raven Progressive Matrices, and meta-cognitive strategies using a battery of 10 questions on planning, evaluation and adjusting oneself learning process. They also measure social abilities, such as leadership, problem solving and team work; and self-efficacy, which refers to the perception about own capacity of the individuals. Each of these measures is also obtained from batteries of 10 questions.

They include the standardized scores from those tests in log earnings regressions. When these measured skills are included in the regression without controlling for education, they find a positive and statistically significant relationship of each skill and labor earnings. The estimated coefficients are larger for non-cognitive skills than for cognitive skills. For instance, their results for males show that a one standard deviation on both the intellectual ability score and the meta-cognitive abilities score is associated to 0.05 higher log earnings. The corresponding figures for auto-efficacy and social ability scores are 0.09 and 0.06 respectively. In addition, the estimated coefficients, except for social ability, are larger for females than for males.

However, when they control for education in the regressions, all the estimated coefficients become smaller and only the coefficients on non-cognitive skills remain statistically significant. As we argue latter, it is likely that measured skills (as opposed to latent skills) are affected by schooling. Therefore, the coefficients of skills when education is not included in the regression partially capture the effect of schooling on earnings.

3. Data and measured skills

Data for this study come from a survey developed specifically to explore the relationship between labor outcomes, education and cognitive and non-cognitive skills in Peru. These data come from a random sample of the working-age (14-50) urban population in Peru (n= 2,660).

The instruments of the survey include typical questionnaire sections of labor surveys, but also a special module on education trajectories and a battery of tests specially designed to measure cognitive skills (Cueto, Baerthel and Muñoz, 2010) and non-cognitive skills or personality traits (Claux and La Rosa, 2010).On the cognitive side, tests include the Peabody Picture Vocabulary Test, and specially designed tests of verbal fluency, working memory, and numeracy/problem-solving. On the non-cognitive side, the survey includes two kinds of batteries to measure personality traits. The first allows measuring the big-five factors (Goldberg 1990) while the second allows measuring the Grit personality traits of perseverance and the will to strive for long term goals (Duckworth et al 2009). Data are also collected on the conditions under which tests are applied to capture variations and measurement error.

The surveys also collects rich information of individual schooling trajectories from pre-school to college on factors known to influence the early acquisition of abilities and access to schooling, such as parental background (mother and father's education, occupations), family structure at young age (number of brothers and sisters, birth order and spacing), distance and quality characteristics of primary and secondary schools and self-reported socioeconomic status while attending, self-reported scholastic performance, effort exerted at school, and parent's valuation of education. This allows us employing usually hard to find possible instrumental variables for schooling and measured abilities in a single dataset.

The working sample used in the study comprises 1,140 observations of male and female working age individuals which were currently employed and working with positive earnings at the time of the survey, for whom complete information on schooling trajectories and tests results are available. Unweighted summary statistics are reported in Table 1.

Measured skills

In the case of cognitive skills, we use standardized scores computed from the original scores on four tests: the Peabody's Picture Vocabulary Test (PPVT), a verbal fluency test, a working memory test, and a Math problem solving test. For the PPVT, working memory and math problem solving scores, we use *Rasch scores* to compute the final standardized scores.[5] Details are provided by Cueto, Muñoz and Baertl (2010).

All our four measures of cognitive skills are partial measures of the individual's intellectual ability and are positively correlated between each other. For these reasons, we also use an aggregate measure of cognitive skills obtained from a principal component analysis

---

[5] This is due to the lack of credible population reference norms for Peru for the PPVT and other cognitive tests. See Cueto et al 2010.

(Cueto, Muñoz and Baertl, 2010). We use the first principal component as our aggregate measure cognitive ability which we interpret as a proxy variable of the individual's intellectual ability.

In the case of non cognitive skills, we consider two broad concepts of personality traits. The first corresponds to the big-five factors of personality according to Goldberg et al (1990). These dimensions are openness, agreeableness, emotional stability, conscientiousness, and extraversion. However, the factor analysis to construct the five factors suggests that a model that splits the agreeableness dimension in two parts best describe the data. One part encompasses traits related kindness, while the other encompasses traits related to cooperation.

The other measures of non-cognitive skills are the Grit personality trait of perseverance and the will to strive for long term goals. Factor analysis leads to two dimensions: consistency of interest and persistence of effort. Claux and La Rosa (2010) provide details on the construction of the five factor model and the Grit personality traits applied to the Peruvian case. We use standardized z-scores obtained from the original scores as our non-cognitive measures.

4. Preliminary evidence on the relationship between earnings and measured skills

We first explore the relationship between log earnings and measured cognitive and non cognitive skills indicators. We have performed a conditional correlation analysis of this relationship by running regressions of log earnings on the cognitive and non-cognitive skill measures partialing out the effect of own schooling, working experience, gender, geographic location (place of residence) and ethnic background, as well as parents' schooling.

In the case of cognitive skills, we find that all our measured cognitive skills indicators are positively related to earnings. However, their estimated impact is mediated by schooling. This is expected as we only have current measures of cognitive skills which might be affected by acquired years of schooling. This issue will be addressed in a latter section of the paper.

Regarding non-cognitive skills, we find a positive relationship between earnings and Goldberg's emotional stability and Grit's persistence of effort, and a negative relationship between earnings and Goldberg's agreeableness-cooperation. The inclusion of schooling in the analysis, however, does not change much the estimated relationship between measured non cognitive skills. This suggests that schooling has a minor o no role on our non-cognitive skills indicators, which are also current measures.

4.1. The relationship between earnings and cognitive skills

A simple graphical analysis shows the positive partial relationship between earnings and our measured cognitive skills indicators. In Figure 1 we display scatter plots of log earnings against the standardized scores of cognitive skills. In all four cases, a positive relationship emerges.

However, these plots do not take into account the effects of other variables in the relationship between measured cognitive skills and earnings. In particular, we explore if the positive relationship remains after controlling for other variables likely related to earnings.

Table 2 we report results from log earnings regressions by including one cognitive score at a time. In columns 1 to 4 we control for the additional covariates except for own schooling. We find positive and statistically significant coefficients for each of our four measured cognitive skill indicators when included one at a time. In particular, a one standard deviation on the PPVT score is associated to 15.8% higher earnings, while one standard deviation on the Math problem solving score is associated to 16.8% higher earnings. Verbal fluency and working memory also have positive and statistically significant coefficients, but their magnitudes are lower: one standard deviation on verbal fluency is associated to 9.8% higher earnings, while working memory increases earnings by 8.9%.

When we further control for the individual's own education, measured in years of completed schooling, in columns 5 to 8, all four measures of cognitive skills remain positive but their size and their statistical significance drops down for three of them. The estimated coefficient on the PPVT score falls and is now barely statistically significant in the regression, with an impact of 6.6% higher earnings for an additional standard deviation on its score. The coefficients on verbal fluency and working memory also drop, with impacts of 3.7% and 3.1% and are no longer statistically significant. The coefficient on math problem solving also drops, with an estimated impact of 10.5%, but remain highly statistically significant. In all cases, the estimated impact of an additional year of schooling varies from 5.7% to 6.8%. These results suggest that part of the correlation between cognitive skills and log earnings materializes trough their effect on years of schooling. Indeed, we find evidence in our sample of a positive correlation between each of our measured cognitive skills indicators and years of schooling. We take this into account and report results latter in the analysis.

All the four measured cognitive indicators we use are partial measures of an individual's intellectual ability and there is a positive correlation between them. An aggregate indicator that combines the four indicators in one single proxy for cognitive ability was constructed using principal components analysis. In Table 3 we report the correlations between the four cognitive indicators as well the correlations of those indicator with the aggregate proxy for cognitive or intellectual ability. The correlation between the four measured cognitive indicators varies from 0.38 (verbal fluency and working memory) to 0.52 (PPVT and math problem solving). All these correlations are highly statistically significant. As expected, the aggregate measure is highly correlated to each of the four measured cognitive indicators with correlation coefficients above 0.73 and also highly statistically significant.

An additional specification we explore includes all the four measured cognitive skills indicators together in the regression. We do this in two ways, by including the four measured cognitive skills indicator in the regression, and also by replacing the individual measured cognitive skills by the aggregate proxy for cognitive ability. In both case we run the regression with and without years of schooling. Results are reported in Table 4.

Including all measured cognitive skills indicators in the regression without own education attenuates the effects of each individual indicator on earnings. The positive relationships remain, but only the PPVT and math scores stay statistically significant. As column 1 from the table shows, one standard deviation on the PPVT has an impact of 9.3% higher earnings, while a standard deviation on the math score has an impact of 12.3%, both below the estimated impacts when included one by one. Controlling for education further attenuates the estimated impacts of measured cognitive skills. As column 3 shows, in this case only the math score remains statistically significant with an estimated impact on earnings of 9.6%. None of the remaining cognitive indicators are statistically significant in this specification.

A positive relationship between cognitive skills and earnings also emerges when including the aggregate cognitive indicator instead of the individual indicators. Our results suggest that intellectual ability contributes to generate higher earnings even after controlling for schooling.

Column 2 from Table 4 shows the regression that includes the aggregate proxy for cognitive ability but without controlling for years of schooling. The estimated coefficient in this regression suggests that an additional standard deviation on the aggregate proxy for cognitive ability increases earnings by 18.2%, and the point estimate is highly statistically significant. When controlling for schooling, as column 4 shows, the estimated impact of our proxy for cognitive ability drops to 10%, but stay highly statistically significant.

There is no a correct or unique way to include the different cognitive skills in the regressions. In particular, the psychometric literature shows that the different dimensions of intellectual ability are not orthogonal to each other. As we showed, all these dimensions of cognitive skills are correlated. Since it is difficult to disentangle the effect of each of these dimensions on earnings, in the rest of the paper we use the aggregate measure of cognitive skills as our preferred indicator for measured intellectual ability. As we mentioned, this aggregate indicator is related to the *g-factor* of intelligence from the psychometric literature.

4.2. The relationship between earnings and non-cognitive skills

Now we turn to measured non-cognitive skills. We explore the relationship between earnings and the big five factors or dimensions of personality according to Goldberg and two Grit personality traits of consistency of interest and perseverance of effort.

As Figure 2 shows, an inspection of scatter plots suggests a positive relationship between earnings and both dimensions of Grit. It also reveals a more complicated relationship emerges between earnings and the big five factors of personality. Extraversion, emotional stability and openness appear to be positively associated to earnings, while agreeableness and conscientiousness appear to be negatively associated to earnings.

Using a regression analysis we partial out the effects of other variables in the relationship between earnings and measured non-cognitive skills just as we did for measured cognitive skills. We find that earnings are positively related to Goldberg's emotional stability as well as to Grit persistence of effort, while negatively related to Goldberg's agreeableness.

In Table 6 we report the results from our log earnings regressions including the two sets of measured non-cognitive skills. The first three columns report regression results that include the five dimensions of personality according to Goldberg and two dimension of long term goal according Grit without controlling for schooling and cognitive skills, while the last three columns include these controls in the regressions.

In the first column we include only the big five dimensions of personality as described by Goldberg. As the results show, log earnings are negatively related to both sub dimensions of agreeableness and positively related to emotional stability. For the agreeableness dimension we find that a one standard deviation on the kindness sub dimension measure is associated to 5.6% lower earnings, but the point estimate is barely significant. In turn, for the emotional stability dimension, a one standard deviation on the measure is associated to 6.7% higher earnings. The result for cooperation sub dimension measure suggests that an increase on this skills dimension is associated to 9% lower earnings. None of the other dimensions of personality has statistically significant coefficients in the regression.

In the second column we include only the Grit measures of consistency of interest and persistence of effort. We find that both are positively related to log earnings, but only the coefficient on persistence of effort is statistically significant. In particular, an increase of one standard deviation on this measure is associated to 12.2% higher earnings when controlling for the additional covariates except own schooling.

Including both Goldberg’s five personality dimensions and Grit personality trait in the regression we find qualitatively similar results regarding the sign of the estimated coefficients, but the statistical significance of the estimated coefficients falls in most cases. The point estimates on the agreeableness-kindness and emotional stability dimensions keep their sign but loss their statistical significance. On the contrary, agreeableness-cooperation and Grit persistence of interest keep their sign, magnitude and statistical significance.

Adding controls for own schooling in the regressions yield similar results. Agreeableness-cooperation and emotional stability dimensions have highly statistically significant coefficients. For agreeableness-cooperation an increase of one standard deviation on the score is associated to 8.2% lower earnings while for emotional stability the estimated effect is associated to 6.7% higher earnings. In the case of Grit, only persistence of effort is statistically significant and implies an 8.9% increase in earnings for a one additional standard deviation in this trait.

In the last column of the table we include both Goldberg's and Grit in the regression that also controls for schooling. Agreeableness-cooperation and emotional stability dimensions as well as Grit persistence of effort remain statistically significant with estimated coefficients similar to those previously estimated.

The results for agreeableness deserve a comment. In this skills dimension, higher scores are associated to more critical, harsh, and rude behavior of the individual; while lower scores are associated to good natured, sympathetic, and forgiving behavior. Thus, the results we find suggest that more critical and harsh behaviors are more valued in the Peruvian labor markets than sympathetic and forgiving behaviors. A recent study by Duckworth and Weir (2010) allows us to place this result in perspective. In their study, they use data from the Health and Retirement Study, which provide information on personality traits, linked to administrative data from the Social Security Administration, which provide well-measured lifetime earnings. They find a negative point estimate for the coefficient of agreeableness in a regression of (log) lifetime Social Security earnings, although their estimate is not statistically significant. They also find a negative and statistically significant relationship between average lifetime earnings and agreeableness. In addition, agreeableness measures for husband and wife among married couples are negatively related with (log) wealth.

5. Econometric models

The traditional earnings regression that researchers in the labor literature aim to estimate is written as:

$$\ln y = \alpha + \beta S + \gamma A + \varepsilon, \tag{1}$$

where ln y are log earnings, S represents acquired years of schooling, and A is ability.

Available data usually do not contain information on ability. Thus, much of the empirical literature has dealt with A as an omitted variable from the true model. The omission of A from the model leads to the problem of omitted ability bias. Depending on the relationship between S and A, the bias may lead to the over or sub estimation of the true parameter β.[6] In particular, if we expect that earnings increase with ability ($\gamma > 0$), and assume a positive relationship between

[6] Heckman, Todd and Lochner (1995) state the conditions that allow interpreting this coefficient as the return to schooling.

ability and schooling, then the OLS estimate of β when ability is omitted from the model will overestimate the true β.

Over the last two decades, a large literature on the estimation of the returns to schooling has developed attempting to solve the omitted ability bias problem in earnings regressions by using the methods of instrumental variables or 2SLS (see Card, 1998 and 2001; Kling 2001). However, much of the evidence that emerged produced larger 2SLS estimates of the returns to schooling instead of lower estimates as one would expect under the overestimation presumption outlined earlier. Summaries of this literature argue that these results are droved by the effect of the specific instrument used on the specific group affected by the instrument (the *compliers*), thus these estimates should be interpreted as local treatment effects instead of average treatment effects.

Another path of research has focused on acquiring and including measures of skills in the earnings regression. At first glance, the inclusion of measured skills should solve the omitted ability bias problem. However, this strategy could be problematic for reasons we describe next.

In line with this venue, we assume that earnings depend on schooling and measured skills (see Altonji and Pierret 2001; Hansen, Heckman and Mullen 2004). Thus, we assume that the true earnings regression can be written as:

$$\ln y = \alpha + \beta S + \gamma T + \varepsilon, \tag{2}$$

where T is a vector that stands for standardized test scores for two kinds of measured skills: cognitive and non-cognitive skills. All these skills are quantified for a random sample of individuals residing in urban areas in Peru at the time the survey used in this study was carried out.

As suggested by Heckman, Stixrud and Urzua (2006), a general model of the effects of skills and schooling on earnings should account for unobserved or latent skills. Latent skills, both cognitive and non-cognitive, affect measured skills (at the time of the survey) and school choice. Measured skills are affected by schooling and family background. Measured skills and schooling affects earnings.

As we discussed in the previous section, the evidence on skill formation suggests that skills evolve over time and it is likely that measured skills are influenced by previously acquired schooling. In particular, since the publication of the work by Herrestein and Murray (1995) on the *Bell Curve*, several studies have presented evidence that point to a positive dependence of measured cognitive skills on schooling (Winship and Korenman 1997, Hansen, Heckman and Mullen 2004).

If this is the case, as we empirically show later, the γ coefficient partially captures the indirect effect of schooling on earnings through the measured skills. Taking derivative of log earnings with respect to schooling we obtain the total effect of schooling on earnings, our parameter of interest for the return to schooling:

$$\frac{\partial \ln y}{\partial S} = \beta + \gamma \frac{\partial T}{\partial S}. \tag{3}$$

As it becomes evident from equation (3), the effect of schooling on earnings cannot be directly obtained from the estimated coefficients from the regression unless we account for the dependence of T on S.

In order to solve this problem, we estimate a two-step model. In the first step we remove the dependence of measured skills on schooling by estimating the following regression:

$$T = \delta_0 + \delta_1 X + \delta_2 S + \eta. \tag{4}$$

We estimate one regression for each of our measured skills. In particular, we assume that the effect of schooling on measured skills is linear for each type of skill. However, given the possibility of reverse causality on this specification as latent skills can affect schooling choices, we estimate this regression by instrumental variables. In the first stage we run a regression of schooling on the same conditioning variables from the T equation plus the instruments for S:

$$S = \pi_0 + \pi_1 X + \pi_2 Z + \nu. \tag{5}$$

The Z variables are our proposed instruments. We use scholastic achievement, effort exerted at school, and time to get to school. These instrumental variables refer to the individual's situation at the time she pursued her highest acquired level of schooling. All these variables are directly related to schooling, but we assume they are only indirectly related to post-schooling measured skills through schooling. We run the T regression by LIML including the instruments one at a time, and then the three instruments together.

We also consider the socioeconomic status of the family at the time of pursuing the highest level of schooling as a potential instrument for schooling. However, we suspect this variable might be a bad instrument as SES likely affects family investments on skill formation. Nevertheless, we include this variable for comparison purposes.

From the second stage, once schooling has been instrumented in equation (4), we obtain the residuals of measured skills, which we denote by $\tilde{T}$. These residuals are orthogonal to schooling by construction.

In the second step of the procedure, we use these residuals in the earnings regression:

$$\ln y = \alpha + \beta S + \gamma \tilde{T} + \varepsilon. \tag{6}$$

We run this last regression by OLS. In this specification, we interpret β as the effect of schooling on earnings and γ as the effect of measured skills on earnings.

A more complicated model, that our data do not allow us to identify, assumes that earnings depend directly on unobserved measures of skills acquired earlier in life. These skills would also affect school choice and measured skills. Heckman, Stixrud and Urzua (2006) develop and estimate a model on of this kind.

This model is much harder to identify because latent skills appear directly on the earnings equation. Having access to instruments for schooling is not enough to identify the model as measured skills remain correlated to the error term in the earnings regression because of the latent skills.

6. Returns to schooling and skills: putting all the pieces together

We now present results of the estimation of the Mincer log earnings regressions augmented with cognitive and non-cognitive skills measures. We first assume that skills and schooling are exogenous variables in the regression. Then we recognize that measured skills might be affected by schooling and implement a procedure to partial out the effect of schooling from the effect of skills on log earnings.

6.1. Results assuming no correlation between measured skills and education

The first column of Table 7 shows the estimated coefficient of years of schooling from a typical log earnings regression in which we control for working experience, gender, geographic location and ethnic background. The point estimate of the schooling coefficient is 0.092 and highly statistically significant, which indicates that earnings increase in 9.2% with an additional year of schooling. However, in this specification of the regression we suspect that the estimated coefficient of own schooling is likely biased because of the potential correlation between schooling and left out variables subsumed in the error term.

In the second column we add controls for parents' schooling. We control for parental education in the regression because part of the correlation between log earnings and own schooling might be explained by an intergenerational transmission of skills between parents and children.[7] In this case, the estimated coefficient of schooling drops 22% to 0.072 and the estimated standard error increases 15%, but nevertheless the point estimate remains highly statistically significant.

When we include the aggregate indicator of cognitive skills and all non cognitive skills indicators in regression (column 3), we find that the return to schooling drops further but the qualitative patterns of the returns to skills previously documented remain almost unchanged. Our estimates suggest that an increase of one standard deviation on the aggregate indicator of cognitive skills is associated to 9.4 percent higher earnings. Increases on emotional stability and Grit's persistence of effort are associated to increases on earnings of 5.7% and 8.3% respectively. On the contrary, an increase of one standard deviation on the agreeableness-cooperation score is associated to 9% lower earnings.

It is worth to mention that the inclusion of measured skills, both cognitive and non-cognitive, generates a reduction of the estimated return to schooling. This suggests that the estimated coefficient of schooling in column 2 overestimates the true effect of schooling on

[7] Some studies on the returns to schooling have used parental background as an instrument for schooling. Even when there is a positive correlation between own schooling and parents' education, it is difficult to argue that parents' education could serve as exclusion restrictions in the log earnings equations. In the most likely scenario, parents' education affects the earnings of their children through other channels in addition to children's own schooling. For this reason, we keep parents' education in the earnings regression equation in our preferred specification, and include further controls that account for potentially additional left out variables in the regression. We do this in column 3, where we further include the cognitive and non cognitive skills measures.

earnings. This result is in line with the literature that postulates a positive relationship between schooling and ability. In this case, when ability is omitted from the regression, the coefficient of schooling is an estimate of both the true effects of schooling on earnings plus the correlation between schooling and ability. It should be stressed that the analysis of this section implicitly assumes that skills are not affected by schooling. However, our skills indicators are current measures of skills that might be indeed affected by schooling. In the next section we allow that measured cognitive and non-cognitive skills are affected by schooling choices.

6.2. Results when measured skills are affected by schooling

A potential problem of the previous estimates is that measured cognitive skills might be affected by schooling choices. All the measured skills in our data were gathered at the time of the survey, therefore they are current measures of cognitive ability. If schooling positively affects measured skills, we are underestimating the effect of schooling in the log earnings regression because part of this effect is captured by the measured cognitive skills in the regression.

In our data, more education seems to be related to higher scores on our aggregate measure of cognitive skills, as well as for extraversion, openness, consistency of interest and persistence of effort. Figure 3 shows density plots of skills measures for people with primary or secondary education (11 or fewer years of schooling) and those with higher education (University and non-University higher education).

From the previous section, a clear pattern emerges from the earnings regressions. When own schooling is included in the regressions, the estimated coefficients of all measured skills drop. The coefficients from regressions that do not include own schooling produce estimates of the net effect of these skills on earnings, that is, they account for the direct effect of skills as well as the effects trough schooling.

However, our measured cognitive skills are not the same as measured intelligence. As many previous studies emphasize (Heckman 1999; Carneiro and Heckman 2003; Carneiro, Heckman and Masterov 2005; Heckman, Cunha, Masterov, Urzua 2006; Heckman, Stixrud and Urzua 2006), while intelligence is fairly well set by the age 8-10, achievement tests scores are malleable over a longer span. In particular, evidence for the US shows that achievement measured by the AFQT score is positively affected by years of schooling (Neal and Johnson

1996, Hansen, Heckman and Mullen 2004). The implication is that including both measured skills and schooling in our OLS earnings regression creates a problem: it is not possible to distinguishing between higher measured ability causing higher wages, from additional years of schooling causing both higher measured skills and higher wages (Hansen, Heckman and Mullen 2004; Heckman, Stixrud and Urzua 2006).

To address these issues, we implement a two-step procedure to remove the effect of schooling on measured skills. In the first step we estimate a series of regressions of our measured skills on schooling and all the other covariates. From these regressions we obtain the residuals which will be included in the log earnings regression. The goal is to purge the effect of schooling on measured skills; we use instrumental variables for schooling so we can capture only the effect of exogenous variation of schooling on the measured skills indicators.

We use three instrumental variables for schooling: scholastic achievement at school; effort exerted at school; and time to get from home to school. These variables are construed from self-report responses to questions on the school trajectories module from the survey. All three instruments are dummy variables constructed from corresponding questions in the survey. Scholastic achievement is a dummy that takes the value one when the individual obtained either good or highest grades and the value zero in other cases. Effort is a dummy that takes the value one when the individual exerted a self-reported high level of effort at school and the value of zero in other cases. Time to school is a dummy that takes the value one when the individual lived half an hour away from school.

In the second step we run the log earnings regression including the residualized skills from obtained from the first step instead of the original measured skills.

Results of the first step are reported in Table 8. Each column of the table reports the result of running an instrumental variables regression of measured skills on years of schooling and additional covariates, where schooling is assumed endogenous.[8] All the estimated coefficients on instrumented schooling are positive in these regressions. Schooling has larger coefficients for

[8] In general, we do not find effects of work experience, sex, and parents' education on the measured skills. However, individuals from Lima (the capital city) have higher scores on our measured skills, except for agreeableness, emotional stability, and consistency of interest. In addition, individual with indigenous ethnic background have lower cognitive skills relative to individuals with European background, while those with African Peruvian background have higher non-cognitive skills.

cognitive ability persistence of interest, and extraversion. For instance, an additional year of schooling increases the aggregate measure of cognitive skills by 0.25 standard deviations, and the measure of persistence of effort by 0.16 standard deviations. The estimated effects of schooling on openness, agreeableness (cooperation), and conscientiousness are smaller. An additional year of schooling increases the measure of openness by 0.11 standard deviations and the measure of agreeableness (cooperation) by 0.07 standard deviations. However, the estimated coefficients of schooling in the agreeableness (kindness), emotional stability, and consistency of interest regressions turned not statistically significant.

Several studies from the U.S. find that schooling increases measured cognitive skills (Neal and Johnson, 1996; Winship and Korenman, 1997; Hansen et al, 2004). All these studies use data from the NSLY and the cognitive skills are measured by the AFQT. Hansen et al use instrumental variables to estimate a regression of AFQT points on instrumented schooling and find that one additional year of schooling increases AFQT by 4.5 points.

To check for the appropriateness of the instrumental variables estimation we compute usual tests of identification and weak instruments from the first-stage regression. We find low values on the Hansen-J test of over identification suggesting that the instruments are valid as they are orthogonal to the error from the structural equation for measured skills. The Wald F test (Kleibergen-Paap) of weak instruments suggests that there are no concerns of weak correlation between schooling and the instruments. The values of the test, 47.3 and 34.8, are well above the corresponding critical values from Stock and Yogo (2005).

Table 9 reports the results of the log earnings regressions using the residualized measures of skills. The first column replicates the full log earnings regression estimated using OLS under the exogeneity assumption. The rest of the columns in the table report the estimated regressions using the two-step procedure. Columns 2 to 4 report results using each instrument for schooling separately, while column 5 reports the results using the three instruments together.

Across these regression results, we find similar qualitatively patterns for the returns to schooling, cognitive and non cognitive skills, although their magnitude changes with the instruments we use. Those patterns are also similar to what we found from the OLS estimation in terms of the signs of the coefficients.

In all of these regressions we find that the point estimates change between specifications, but the magnitude of the change is not large enough to make them different in a statistical sense. For instance, the estimated return to schooling increases to 0.055 when we use the scholastic achievement instrument alone and the three instruments together (columns 2 and 5) compared to the point estimate of 0.048 from OLS. When we use the effort and time instruments, however, the point estimates from 2SLS are 0.048 and 0.047 respectively, not different from the OLS result. Something similar happens for the other coefficients.

We also consider SES of the family at the time the individual was pursuing education as a potential instrument for schooling. This variable is included as the SES at the time of pursuing primary education, and at the time of pursuing the highest acquired level of education. Since SES likely also affects investments on skill formation, it is more difficult to assume that SES as an exclusion restriction. Even more, we find evidence that suggests a problem of weak instruments. Using the test of weak instruments we find an F statistic of 8 when we use the indicator of SES at the primary level, and an F statistic of 13.1 when we use the indicator of SES at the highest education level. These estimates lie below the critical values of Stock and Yogo.

Nevertheless, for comparison purposes we also estimate the earnings regressions in the second step when SES was used as an instrument for schooling in the first step. Results are reported in Table 10. The estimated coefficients are similar to those reported in Table 9, with the difference that the estimate for the return to schooling is a bit lower and that the coefficient of persistence of effort remain statistically significant.

As a wrap-up, we find that both schooling and measured skills are valued in the Peruvian labor market. This result is consistent both under independence between measured skills and schooling, and under the presumption that schooling affects measured skills.

In the first case, our OLS results suggests that increasing years of schooling by one standard deviation (which corresponds to an increase of about 3 years in our working sample) increases hourly earnings by 14%, increasing the score of cognitive skills by the same magnitude increases hourly earnings by 9.4%, increasing the score of emotional stability increases hourly earnings by 5.7%, and increasing the score of persistence of effort increases earnings by 8.3%. On the contrary, increasing agreeableness (cooperation) reduces earnings by 9%. As we mentioned, this might appear an awkward result, but Duckworth and Weir (2010) find a similar association between earnings and agreeableness using U.S. earnings data.

In the second case, when we allow that schooling affects measured skills, our 2SLS results are similar, but the effect of schooling on earnings is a bit larger while the effects of skills are a bit smaller. In this case, an increase of schooling by a one standard deviation increases earnings between 14% and 16.5%, while an increase on cognitive skills increases earnings between 8.2% and 10%. The effect of agreeableness and emotional stability remain almost unchanged.

7. Conclusions

The evidence presented in this paper shows that conditional on education; both cognitive and non-cognitive measured skills pay off in the Peruvian labor market. The effect of measured cognitive skills, net of schooling effects, is positive and statistically significant. Non-cognitive skills are also positively valued in the Peruvian labor market. There are many dimensions of personality that cannot be summarized in a single factor. Because of this, we have used the big five factors model of personality traits and the Grit personality trait to study the effects of non-cognitive measured skills on earnings.

We corroborate findings from developed countries that both cognitive and non-cognitive skills are important correlates of earnings. After correcting for the potential endogeneity of measured skills vis-à-vis schooling, the findings confirm that both socio-emotional and cognitive skills are equally valued in the Peruvian labor market. A one standard deviation change in an overall cognitive skill measure and in the perseverance facet of Grit each generates a 9% increase on average earnings, conditional on schooling. The effect size of an increase in years of schooling (about 3 years) is a 15% increase in earnings, conditional on skills. The returns to other socio-emotional skills vary across dimensions of personality: 5% higher earnings for emotional stability while 8% lower earnings for agreeableness (a finding replicated in the U.S by Duckworth et al 2011; also see Roberts et al 2011).

The results are partially aligned with most of the qualities Peruvian employers value (see World Bank 2011). However, some of those qualities (e.g., responsibility, tidiness) would relate directly to the trait of Conscientiousness. Yet no significant relations between this trait and earnings were found. Moreover, Agreeableness linked to cooperation correlates with lower earnings. Yet employers seem to value "interpersonal skills" that would be correlated with agreeableness.

There are some possible reasons for these discrepancies, including limitations of the broad Big-five personality trait data to proxy for narrower socio-emotional skills. First, it is well known in the personality psychology literature that responses on self-reported scales are affected by 'social desirability bias', that is, people may tend to respond more according to how they would like to be seen by others rather than by how they actually behave regularly. A closer examination of the data suggests that the responses to the items on Conscientiousness maybe more severely affected by this problem. The responses are skewed toward positive self-assessments so that the range of variation of the data in the sample is very limited. This could explain the insignificant results obtained in the earnings analysis.

In the case of agreeableness, there are at least two possible interpretations of the results. It may be that although employers value cooperation for keeping a good team environment, in reality less cooperative people are more likely to get ahead by doing better than others rather than cooperating. On the other hand, it may be that at the low end of the distribution, being highly "agreeable" might lead to extreme passivity or represent lack of assertiveness or initiative, which might result in lower wage levels. Anecdotal evidence suggests this is a plausible phenomenon in Peru's labor market. These are issues that warrant further research. In particular, assertiveness skills linked to problem solving and decision-making may be important to consider within the social-emotional skill framework. It would be important for future studies to examine more refined constructs of skills to complement measures of broad personality traits.

The cognitive and socio-emotional skills that have been measured do not exhaust the mechanisms by which schooling affects earnings (and thus labor productivity). The average return to schooling remains significant at nearly 7 percent, reduced by roughly 2.5 percentage points after we account for measured skills. This is a fall from 9.6 percent when we do not control for skill measures. This finding is consistent with other international studies and suggests that a significant portion of the returns to schooling reflects that schooling goes hand in hand with the development of generic skills, but that the lion share is due to aspects or correlates of schooling largely unrelated to the skills measured. That is, the more educated Peruvian workers earn more not solely because schooling proxies for those with higher innate ability, better parental social status and other traits rewarded in the labor market.

The fact that both cognitive and non-cognitive dimensions of skills have are valued in the Peruvian labor market beyond their potential impact trough the education channel (more able individuals acquire more schooling), have important policy implications. The first is the call to emphasize investments in early childhood as cognitive and non cognitive skills begin to develop early in life. Contrary to cognitive skills, non cognitive skills remain malleable over youth and adulthood. Thus, interventions to increase this dimension of ability should have positive returns.

This calls for a policy framework that goes beyond narrow and fragmented educational, training and labor policies and integrates them into a long term skills development strategy that—starting from investments in early-childhood development of poor children—strengthen degree completion and schooling transitions, improve education quality, and the functioning of training and labor markets in Peru.

References

Altonji, J., and C. Pierret, (2001). "Employer learning and statistical discrimination." *Quarterly Journal of Economics*, vol. 116, no. 1, pp. 313-350.

Bassi, M. and S. Galiani. (2009). Labor Market Insertion of Young Adults in Chile. Inter-American Development Bank, unpublished manuscript.

Ben-Porath, Y. (1967). "The production of human capital and the life cycle of earnings." *Journal of Political Economy*, vol. 75, no. 4, Part 1, pp. 352-365.

Borghans, L., A. Duckworth, J. Heckman and B. ter Weel (2008). "The Economics and Psychology of Personality Traits." *Journal of Human Resources*, vol. 43, no. 4, pp. 972-1059.

Bowles, S., H. Gintis, and M. Osborne. (2001). "The determinants of earnings: A behavioral approach." *Journal of Economic Literature*, vol. 39, no. 4, pp. 1137-1176.

Card, D. (1998). "The Causal Effect of Education on Earnings." Ashenfelter, O. and D. Card (Eds.), *Handbook of Labor Economics*, vol. 3A, pp. 1801-1863.

Card, D. (2001). "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, vol. 69, no. 5, pp. 1127-1160.

Carneiro, P., and J. Heckman. (2003). "Human capital policy." In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press, pp. 77-239.

Carneiro, P., J. Heckman, and E. Vytlacil. (2006). "Estimating marginal and average returns to education." Unpublished manuscript. University of Chicago, Department of Economics.

Claux, M. and M. La Rosa. (2010). "Estudio de factores relacionados con la empleabilidad en zonas urbanas del Perú. Desarrollo de escalas de personalidad y emprendimiento." Unpublished manuscript.

Cueto, S., I. Muñoz, and A. Baertl. (2010). "Scholastic Achievement, Cognitive Skills and Personality Traits of Youths and Adults in Peru: A cross-sectional and intergenerational analysis." Unpublished manuscript. GRADE, Lima.

Cunha, F., and J. Heckman. (2007). "The technology of skill formation." *American Economic Review*, vol. 97, no. 2, pp. 31-47.

Cunha, F. and J. Heckman. (2008). "A New Framework for the Analysis of Inequality." *Macroeconomic Dynamics*, vol. 12, no. S2, pp. 315-354.

Cunha, F., and J. Heckman. (2008). "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources*, vol. 43, no. 4, pp. 738-782.

Cunha, F., and J. Heckman. (2009). "The economics and psychology of inequality and human development." *Journal of the European Economic Association*, vol. 7, no. 2-3, pp. 320-364.

Cunha, F., J. Heckman, L. Lochner, and D. Masterov. (2006). "Interpreting the evidence on life cycle skill formation." In E. A. Hanushek, and F. Welch (Eds.), *Handbook of the Economics of Education*, ch. 12, pp. 697-812. Amsterdam: North-Holland.

Cunha, F., J. Heckman, and S. Schennach. (2010). „Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, vol. 78, no. 3, pp. 883-931.

Goldberg, L. (1993). "The structure of phenotypic personality traits." *American Psychologist*, vol. 48, no. 1, pp. 26-34.

Duckworth, A., and D. Weir. (2010). "Personality, Lifetime Earnings, and Retirement Wealth." Unpublished manuscript.

Hansen, K., J. Heckman, and K. Mullen. (2004). "The effect of schooling and ability on achievement test scores." *Journal of Econometrics*, vol. 121, no. 1-2, pp. 39-98.

Heckman, J. (1999). "Policies to Foster Human Capital." NBER Working Paper, no. 7299.

Heckman, J., J. Stixrud and S. Urzua (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior". *Journal of Labor Economics*, vol. 24, no. 3, pp. 411-482.

Heckman, J., and E. Vytlacil. (2001). "Identifying the role of cognitive ability in explaining the level of and change in the return to schooling." *Review of Economics and Statistics*, vol. 83, no. 1, pp. 1-12.

Herrnstein, R., and C. Murray. (1994). *The Bell Curve: Intelligence and class structure in American life*. New York: Free Press.

Kling, J. (2001). "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business & Economic Statistics*, vol. 19, pp. 358–364.

Mincer, J. (1958). *Schooling, Experience, and Earnings*. Columbia University Press, New York.

Murnane, R., J. Willett, and F. Levy. (1995). "The growing importance of cognitive skills in wage determination." *Review of Economics and Statistics*, vol. 77, no. 2, pp. 251-266.

Neal, D., and W. Johnson. (1996). "The role of premarket factors in black-white wage differences." *Journal of Political Economy*, vol. 104, no. 5, pp. 869-895.

Roberts, B., Jackson, J.J., Duckworth, A., and K. Von Culin. (2011). Personality Measurement and Assessment in Large Panel Surveys. Forum for Health Economics and Policy, 14(3).

Roberts, B. W., Kuncel, N., Shiner, R., N., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. Perspectives in Psychological Science, 2, 313-345.

Shonkoff, J., and D. Phillips. (2000). *From neurons to neighborhoods: The science of early child development*. Washington, DC: National Academy Press.

Winship, C., and S. Korenman. (1997). "Does staying in school make you smarter? The effect of education on IQ in the bell curve." In: Devlin, B., S. Fienberg, D. Resnick, and K. Roeder

(Eds.), *Intelligence, Genes, and Success: Scientists respond to The Bell Curve*. Copernicus Press, New York, pp. 215-234.

Table 1
Summary statistics

	N	Mean	Std. Dev.	Min.	Max.
Hourly earnings (logs)	1140	1.223	0.873	-3.258	5.154
Years of schooling	1140	11.422	3.164	1.000	19.000
Cognitive skills					
PPVT	1140	0.115	0.987	-3.195	2.760
Verbal fluency	1140	0.043	1.013	-2.565	4.686
Working memory	1140	-0.010	1.004	-2.960	3.584
Math problem solving	1140	-0.003	1.019	-3.375	2.339
Aggregate cognitive measure	1140	0.042	1.019	-2.779	3.205
Non-cognitive skills					
Goldberg: extraversion	1140	0.105	0.979	-3.236	1.762
Goldberg: agreeableness, kindness	1140	0.029	0.976	-4.412	0.892
Goldberg: agreeableness, cooperation	1140	0.040	1.016	-4.537	1.180
Goldberg: conscientiousness strong	1140	0.084	0.978	-3.785	1.111
Goldberg: emotional stability	1140	0.078	0.990	-3.277	1.552
Goldberg: openness	1140	0.103	0.990	-3.929	1.496
GRIT 2 Consistency of interest	1140	-0.033	1.019	-2.990	1.956
GRIT 2 Persistence of effort	1140	0.189	0.947	-3.320	1.685
Work experience	1140	1.378	0.965	0	4.100
Work experience squared (x100)	1140	2.829	3.166	0	16.810
Sex (male =1)	1140	0.504	0.500	0	1
Residence: Lima	1140	0.262	0.440	0	1
Residence: Jungle	1140	0.218	0.413	0	1
Residence: Highlands	1140	0.239	0.427	0	1
Ethnic background: Quechua	1140	0.121	0.326	0	1
Ethnic background: Other native	1140	0.030	0.170	0	1
Ethnic background: White	1140	0.057	0.232	0	1
Ethnic background: Afro Peruvian	1140	0.014	0.118	0	1
Ethnic background: Other	1140	0.023	0.149	0	1
Father education: Elementary	1140	0.374	0.484	0	1
Father education: High school	1140	0.325	0.469	0	1
Father education: Tertiary	1140	0.165	0.371	0	1
Father education: Unknown	1140	0.057	0.232	0	1
Mother education: Elementary	1140	0.382	0.486	0	1
Mother education: High school	1140	0.279	0.449	0	1
Mother education: Tertiary	1140	0.124	0.329	0	1
Mother education: Unknown	1140	0.037	0.188	0	1

Notes: Unweighted statistics.
Source: Encuesta Nacional de Habilidades, ENHAB. The World Bank.

Table 2
OLS estimates of returns to cognitive skills
Measured skills included individually

	(1) Without own schooling in the regression	(2) With own schooling in the regression
PPVT	0.158***	0.066*
	[0.028]	[0.033]
Verbal fluency	0.098**	0.037
	[0.037]	[0.035]
Working memory	0.089***	0.031
	[0.025]	[0.019]
Math problem solving	0.168***	0.105***
	[0.035]	[0.031]

Notes: Each cell reports the coefficient estimated from a separate regression. The dependent variable is the log of hourly earnings. All regressions control for work experience, gender, ethnic background, geographic location, and parental education. Regressions in column (2) also control for schooling.
Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 3
Correlation between cognitive skills indicator and the aggregate cognitive indicator

	Cognitive skills indicators				
	PPVT	Verbal Fluency	Working memory	Numeracy	Aggregate measure
PPVT	1.000				
Verbal fluency	0.442	1.000			
Working memory	0.456	0.383	1.000		
Math problem solving	0.520	0.447	0.442	1.000	
Aggregate cognitive measure	0.802	0.729	0.725	0.804	1.000

Note: All the partial correlations reported in the table are statistically significant at the 0.001 level.

Table 4
OLS estimates of returns to cognitive skills: including all the indicators together

	(1)	(2)	(3)	(4)
	Without own schooling in the regression		With own schooling in the regression	
PPVT	0.093**		0.036	
	[0.035]		[0.041]	
Verbal fluency	0.015		0.002	
	[0.034]		[0.033]	
Working memory	0.008		-0.003	
	[0.030]		[0.029]	
Math problem solving	0.123***		0.096***	
	[0.038]		[0.034]	
Aggregate cognitive measure		0.182***		0.100***
		[0.033]		[0.028]
Years of schooling			0.052***	0.054***
			[0.015]	[0.014]
Observations	1140	1140	1140	1140
R-squared	0.163	0.157	0.181	0.177

Notes: Each column reports the result from a separate regression. The dependent variable is the log of hourly earnings. All regressions control for work experience, gender, ethnic background, geographic location, and parental education.
Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 5
OLS estimates of returns to non-cognitive skills
Measured non-cognitive skills included individually

	(1) Without own schooling in the regression	(2) With own schooling in the regression
Goldberg: extraversion	0.045	0.045
	[0.042]	[0.042]
Goldberg: agreeableness, kindness	-0.033	-0.045
	[0.031]	[0.029]
Goldberg: agreeableness, cooperation	-0.048*	-0.059**
	[0.024]	[0.022]
Goldberg: conscientiousness strong	-0.008	-0.023
	[0.026]	[0.025]
Goldberg: emotional stability	0.074***	0.052**
	[0.022]	[0.022]
Goldberg: openness	0.069**	0.024
	[0.031]	[0.029]
GRIT 2 Consistency of interest	0.018	-0.002
	[0.034]	[0.032]
GRIT 2 Persistence of effort	0.122***	0.089**
	[0.036]	[0.039]

Notes: Each cell reports the coefficient estimated from a separate regression. The dependent variable is the log of hourly earnings. All regressions control for work experience, gender, ethnic background, geographic location, and parental education. Regressions in column (2) also control for schooling.

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 6
OLS estimates of returns to non-cognitive skills included by groups

	(1)	(2)	(3)	(4)	(5)	(6)
	Without own schooling in the regression			With own schooling in the regression		
Goldberg: extraversion	0.089		0.077	0.067		0.058
	[0.054]		[0.053]	[0.058]		[0.057]
Goldberg: agreeableness, kindness	-0.056*		-0.048	-0.050		-0.045
	[0.032]		[0.034]	[0.032]		[0.032]
Goldberg: agreeableness, cooperation	-0.090***		-0.087***	-0.082**		-0.080**
	[0.030]		[0.031]	[0.030]		[0.031]
Goldberg: conscientiousness strong	-0.007		-0.034	-0.007		-0.027
	[0.033]		[0.038]	[0.031]		[0.037]
Goldberg: emotional stability	0.067***		0.043	0.067**		0.049*
	[0.024]		[0.027]	[0.025]		[0.027]
Goldberg: openness	0.052		0.031	0.015		0.000
	[0.045]		[0.042]	[0.043]		[0.040]
GRIT 2 Consistency of interest		0.019	0.015		0.000	-0.000
		[0.032]	[0.036]		[0.031]	[0.034]
GRIT 2 Persistence of effort		0.122***	0.113**		0.089**	0.091*
		[0.036]	[0.046]		[0.039]	[0.049]
Years of schooling				0.066***	0.066***	0.063***
				[0.016]	[0.015]	[0.016]
Observations	1140	1140	1140	1140	1140	1140
R-squared	0.147	0.138	0.159	0.186	0.178	0.194

Notes: Each column reports the result from a separate regression. The dependent variable is the log of hourly earnings. All regressions control for work experience, gender, ethnic background, geographic location, and parental education.
Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 7
OLS estimates of the returns to schooling, cognitive and non-cognitive skills

	(1)	(2)	(3)
	No parental background	With parental background	
		Without Skills	With all Skills
Years of schooling	0.092***	0.072***	0.048***
	[0.012]	[0.014]	[0.016]
Cognitive skills			
Aggregate cognitive measure			0.094***
			[0.026]
Non cognitive skills			
Goldberg: extraversion			0.052
			[0.058]
Goldberg: agreeableness, kindness			-0.040
			[0.033]
Goldberg: agreeableness, cooperation			-0.090***
			[0.031]
Goldberg: conscientiousness strong			-0.017
			[0.036]
Goldberg: emotional stability			0.057**
			[0.027]
Goldberg: openness			-0.011
			[0.039]
GRIT 2 Consistency of interest			-0.003
			[0.034]
GRIT 2 Persistence of effort			0.083*
			[0.048]
Observations	1140	1140	1140
R-squared	0.149	0.169	0.201

Notes: Each column reports the result from a separate regression. The dependent variable is the log of hourly earnings. All regressions control for work experience, gender, ethnic background, and geographic location. Except for column (1) all regressions also control for parental background.
Standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8
Regressions of measured skills on schooling

Dependent variable:	(1) Aggregate cognitive measure	(2) Extraversion	(3) Agreeable-ness, kindness	(4) Agreeable-ness, cooperation	(5) Conscien-tiousness	(6) Emotional stability	(7) Openness	(8) Consistency of interest	(9) Persistence of effort
IV for Schooling:	Scholastic achievement, effort, time to school								
Years of schooling	0.250***	0.134***	0.019	0.071*	0.059**	0.049	0.112***	0.035	0.158***
	[0.030]	[0.038]	[0.033]	[0.038]	[0.029]	[0.031]	[0.038]	[0.051]	[0.037]
Work exp.	0.094	0.226	0.132	-0.018	0.243*	0.147	0.109	0.033	0.093
	[0.134]	[0.162]	[0.182]	[0.114]	[0.131]	[0.155]	[0.165]	[0.132]	[0.158]
Work exp2 (x100)	-0.031	-0.033	-0.013	0.033	-0.040	-0.034	-0.011	-0.006	-0.019
	[0.038]	[0.044]	[0.049]	[0.034]	[0.037]	[0.047]	[0.045]	[0.043]	[0.048]
Sex	0.104	-0.028	-0.024	-0.083	-0.096	0.102	-0.013	-0.107	0.051
	[0.061]	[0.077]	[0.085]	[0.095]	[0.076]	[0.090]	[0.073]	[0.076]	[0.076]
Lima Metropolitan	0.202*	0.217**	-0.041	0.113	0.282***	0.095	0.316**	0.036	0.336**
	[0.110]	[0.096]	[0.061]	[0.104]	[0.099]	[0.114]	[0.151]	[0.088]	[0.124]
Jungle	-0.016	0.277	-0.105	0.041	0.193	-0.029	0.232	0.385***	0.193
	[0.110]	[0.209]	[0.098]	[0.149]	[0.170]	[0.137]	[0.181]	[0.134]	[0.202]
Highlands	-0.101	0.036	-0.171**	-0.271	0.005	-0.263*	0.124	-0.161	0.103
	[0.135]	[0.119]	[0.069]	[0.169]	[0.134]	[0.153]	[0.160]	[0.172]	[0.128]
Quechua	-0.363**	-0.022	-0.003	-0.120	0.071	0.154	-0.144	-0.000	0.096
	[0.137]	[0.166]	[0.132]	[0.131]	[0.115]	[0.120]	[0.162]	[0.196]	[0.179]
Other native	-0.247*	-0.260	0.079	-0.108	0.046	-0.334	-0.456*	-0.165	-0.056
	[0.142]	[0.364]	[0.185]	[0.171]	[0.200]	[0.291]	[0.238]	[0.118]	[0.199]
White	-0.236*	-0.301	-0.001	-0.139	0.044	-0.037	-0.192	-0.169	-0.317
	[0.133]	[0.206]	[0.135]	[0.204]	[0.168]	[0.167]	[0.169]	[0.165]	[0.229]
Afro Peruvian	0.041	0.676***	0.370**	0.466**	0.389*	0.404**	0.663***	-0.043	0.610**
	[0.192]	[0.182]	[0.181]	[0.171]	[0.209]	[0.191]	[0.183]	[0.344]	[0.230]
Other	-0.373	-0.433*	0.120	-0.431*	0.316*	0.088	0.179	0.001	-0.240
	[0.222]	[0.218]	[0.304]	[0.245]	[0.177]	[0.179]	[0.121]	[0.264]	[0.223]

Father: Elementary	-0.124	-0.005	0.076	-0.002	-0.026	-0.041	-0.206	-0.037	-0.384**
	[0.143]	[0.196]	[0.174]	[0.240]	[0.159]	[0.241]	[0.198]	[0.247]	[0.152]
Father: High school	0.046	-0.035	0.098	0.013	-0.015	-0.019	-0.090	-0.091	-0.493**
	[0.133]	[0.212]	[0.169]	[0.273]	[0.202]	[0.264]	[0.226]	[0.271]	[0.218]
Father: Tertiary	0.049	-0.124	0.084	-0.108	-0.171	-0.024	-0.066	-0.247	-0.471*
	[0.213]	[0.233]	[0.199]	[0.276]	[0.238]	[0.277]	[0.270]	[0.323]	[0.252]
Father: Unknown	-0.325	0.131	0.145	0.185	-0.070	-0.225	-0.250	0.211	-0.502**
	[0.201]	[0.283]	[0.206]	[0.300]	[0.301]	[0.346]	[0.307]	[0.413]	[0.220]
Mother: Elementary	-0.043	-0.147	0.110	-0.019	-0.131	0.082	-0.240	0.029	-0.063
	[0.111]	[0.131]	[0.122]	[0.148]	[0.145]	[0.119]	[0.150]	[0.145]	[0.158]
Mother: High school	-0.039	-0.131	0.091	-0.026	-0.142	0.203	-0.275	0.108	-0.121
	[0.107]	[0.174]	[0.168]	[0.197]	[0.165]	[0.169]	[0.179]	[0.199]	[0.196]
Mother: Tertiary	-0.104	-0.008	0.074	-0.070	-0.252	0.015	-0.432*	0.483**	-0.304
	[0.191]	[0.183]	[0.190]	[0.245]	[0.203]	[0.180]	[0.243]	[0.191]	[0.246]
Mother: Unknown	0.085	-0.405	0.105	-0.256	-0.211	-0.113	-0.200	-0.233	-0.304
	[0.308]	[0.261]	[0.171]	[0.239]	[0.413]	[0.322]	[0.369]	[0.332]	[0.321]
Constant	-2.831***	-1.603***	-0.384	-0.717*	-0.739**	-0.651	-1.085***	-0.483	-1.320***
	[0.275]	[0.350]	[0.349]	[0.397]	[0.333]	[0.404]	[0.363]	[0.421]	[0.354]
Observations	1140	1140	1140	1140	1140	1140	1140	1140	1140
R-squared	0.445	0.071	0.029	0.040	0.046	0.063	0.104	0.045	0.017

Note: Instrumental variable regressions estimated using LIML. The instruments for schooling are: scholastic achievement, effort exerted, and time to school.

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 9
Second step estimates of the returns to schooling, cognitive and non-cognitive skills

	(1)	(2)	(3)	(4)	(5)
Method:	OLS	Two-Step	Two-Step	Two-Step	Two-Step
Skills measures:	Original	Residualized	Residualized	Residualized	Residualized
IV for Schooling:		Achievement	Effort	Time	All
Years of schooling	0.048***	0.055***	0.048***	0.047***	0.055***
	[0.016]	[0.015]	[0.016]	[0.016]	[0.016]
Cognitive measures					
Aggregate cognitive measure	0.094***	0.082***	0.100***	0.095***	0.088***
	[0.026]	[0.028]	[0.024]	[0.027]	[0.027]
Non cognitive measures					
Goldberg: extraversion	0.052	0.051	0.054	0.060	0.052
	[0.058]	[0.058]	[0.059]	[0.059]	[0.059]
Goldberg: agreeableness, kindness	-0.040	-0.040	-0.041	-0.043	-0.041
	[0.033]	[0.032]	[0.032]	[0.033]	[0.032]
Goldberg: agreeableness, cooperation	-0.090***	-0.088***	-0.087***	-0.088***	-0.088***
	[0.031]	[0.030]	[0.029]	[0.030]	[0.030]
Goldberg: conscientiousness strong	-0.017	-0.018	-0.016	-0.018	-0.018
	[0.036]	[0.036]	[0.036]	[0.037]	[0.036]
Goldberg: emotional stability	0.057**	0.057**	0.058**	0.059**	0.057*
	[0.027]	[0.028]	[0.028]	[0.028]	[0.028]
Goldberg: openness	-0.011	-0.013	-0.013	-0.012	-0.013
	[0.039]	[0.040]	[0.039]	[0.039]	[0.040]
GRIT 2 Consistency of interest	-0.003	-0.004	-0.002	-0.005	-0.004
	[0.034]	[0.034]	[0.034]	[0.034]	[0.034]
GRIT 2 Persistence of effort	0.083*	0.082	0.084*	0.085*	0.084
	[0.048]	[0.050]	[0.048]	[0.049]	[0.051]
Observations	1140	1140	1140	1140	1140
R-squared	0.201	0.199	0.203	0.203	0.200

Notes: Each column reports the result from a separate regression. The dependent variable is the log of hourly earnings. Columns (2) to (5) report estimated coefficients from earnings regressions that control for residualized measured skills. These residuals were obtained from instrumental variable regressions of each skill on schooling and the other covariates.

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 10
Alternative second step estimates of the returns to schooling, cognitive and non-cognitive skills using SES as IV for schooling in the first step

	(1)	(2)	(3)
Method:	OLS	Two-Step	Two-Step
Skills measures:	Original	Residualized	Residualized
IV for Schooling:		SES at primary	SES at highest level
Years of schooling	0.048***	0.049***	0.051***
	[0.016]	[0.016]	[0.016]
Cognitive measures			
Aggregate cognitive measure	0.094***	0.088***	0.081***
	[0.026]	[0.027]	[0.027]
Non cognitive measures			
Goldberg: extraversion	0.052	0.049	0.046
	[0.058]	[0.058]	[0.058]
Goldberg: agreeableness, kindness	-0.040	-0.041	-0.042
	[0.033]	[0.032]	[0.032]
Goldberg: agreeableness, cooperation	-0.090***	-0.090***	-0.088***
	[0.031]	[0.030]	[0.031]
Goldberg: conscientiousness strong	-0.017	-0.017	-0.017
	[0.036]	[0.036]	[0.036]
Goldberg: emotional stability	0.057**	0.056*	0.054*
	[0.027]	[0.028]	[0.028]
Goldberg: openness	-0.011	-0.009	-0.008
	[0.039]	[0.039]	[0.039]
GRIT 2 Consistency of interest	-0.003	-0.003	-0.003
	[0.034]	[0.034]	[0.034]
GRIT 2 Persistence of effort	0.083*	0.084*	0.085*
	[0.048]	[0.048]	[0.048]
Observations	1140	1140	1140
R-squared	0.201	0.200	0.199

Notes: Columns (2) and (3) report estimated coefficients from earnings regressions that control for residualized measured skills. These residuals were obtained from instrumental variable regressions of each skill on schooling and the other covariates.

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Figure 1
Log earnings vs. measured cognitive skills

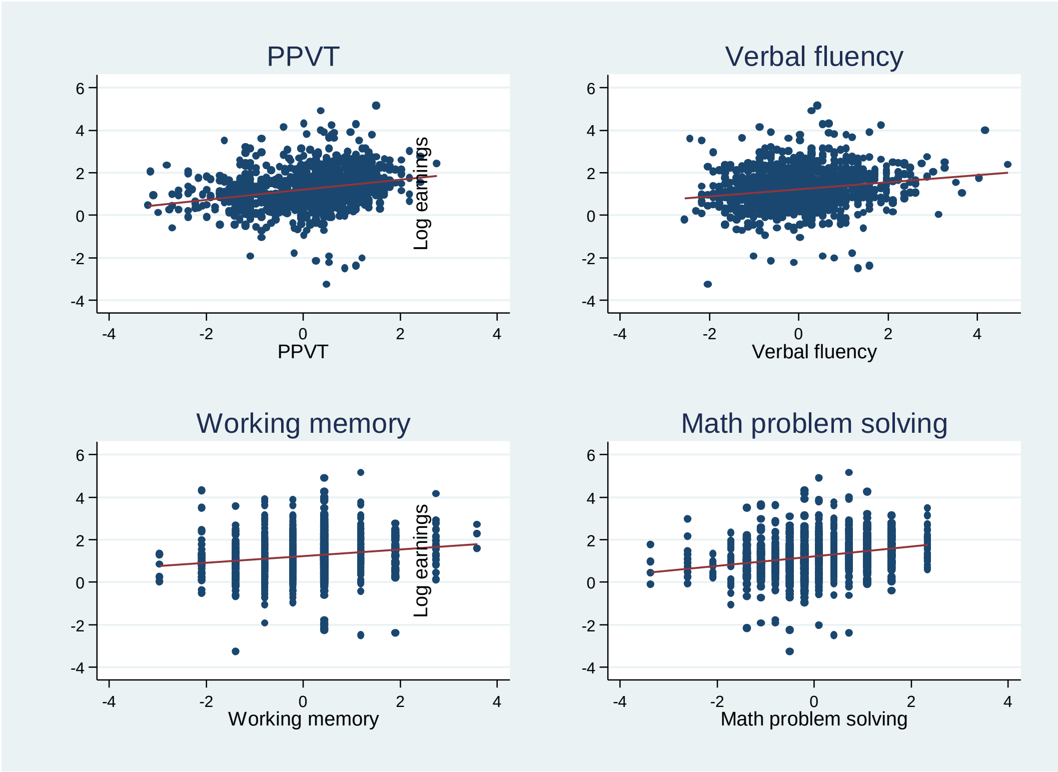


Figure 2
Log earnings vs. measured non cognitive skills

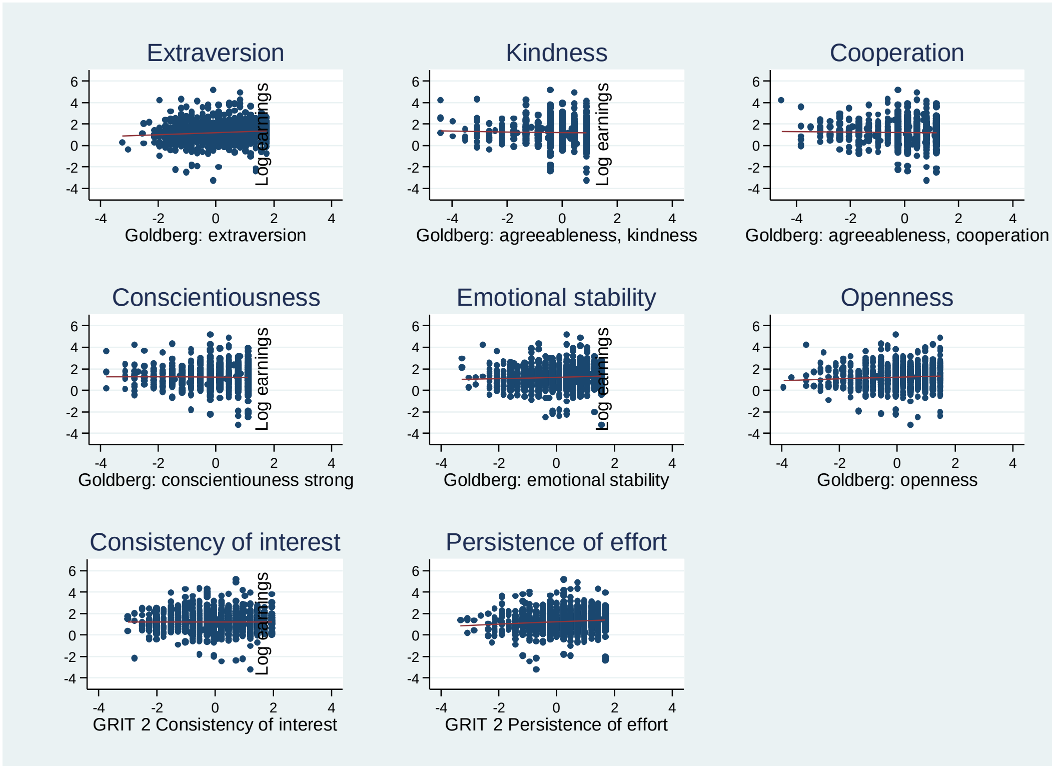
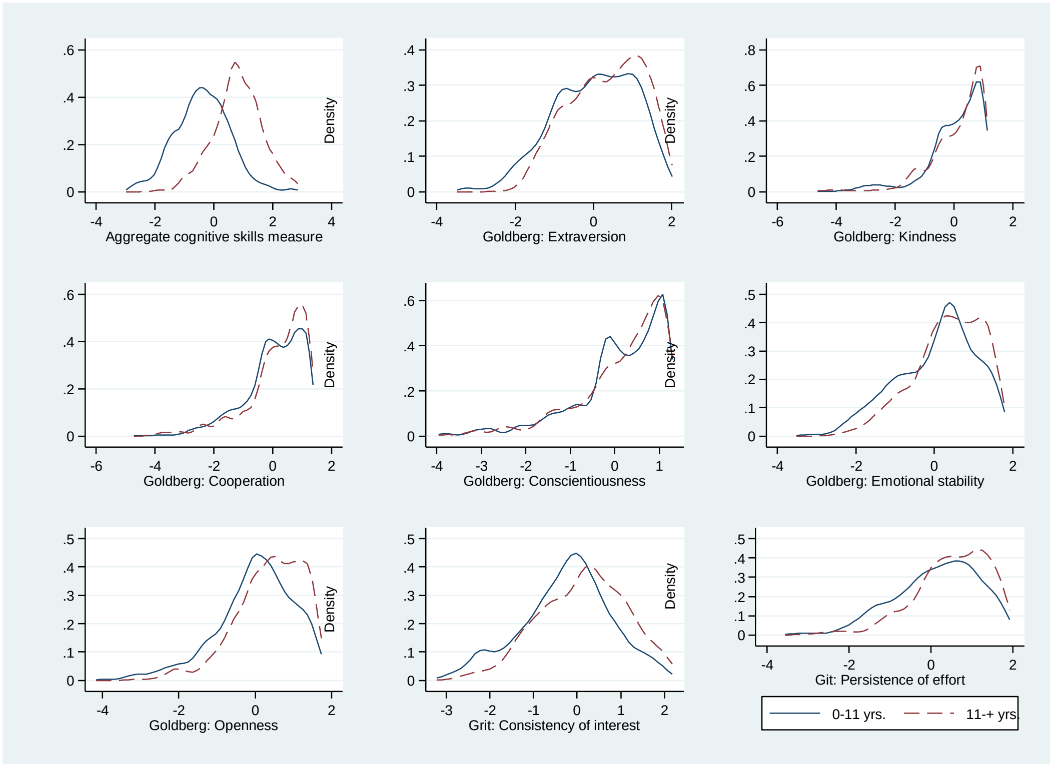


Figure 3
Skills by education level



ANNEX A. Peru Skills and Labor Market Survey

- Self-standing nationally representative of urban areas (2,666 households from cities population >70k), Coast, Highland, Jungle, and Metropolitan Lima.
- Instrument built on existing household survey (ENAHO), housing living conditions, demographics, educational attainment, employment/income module (almost identical), supplemented by:
  - Cognitive and non-cognitive skills tests
  - Labor insertion (first job, tenure, job search, skills certification, reservation wages, mobility disposition, self-employment preferences)
  - School trajectories initial through college/technical education (access, quality proxies, self-reported aptitudes, parental involvement, family conditions, choice of career and institution and reasons, short-term training)
  - Family background (parental education, occupation), relation to siblings (number, birth order variables)
  - Developed over 1+ year (2 pilots). Data collection Jan-March 2010

Skills Measurement

- <u>Sample</u>: age 14-50, one randomly-chosen (pre-field) member per HH (n= 2,666) without replacement (exclude illiterate, non-Spanish speaker)
- <u>Cognitive tests</u> (after pilot validation/revisions):
  - PPVT 4 (verbal perceptive ability, images are shown and must be matched to words, standardized protocol)
  - Verbal fluency (# valid P-words in 3 minutes)
  - Short-term Memory (ability to recall progressive sequence of digits read to test taker)
  - Numeracy-problem solving (18-item multiple choice test, timed 15 minutes)
  - <u>Personality tests</u>
  - BFF 35-item bipolar adjectives, short-sentenced inventory (pre-tested in Lima student population) and 17-item GRIT scale (adapted to Peruvian context)
  - Special, intensified training and evaluation of enumerators (chose best).
  - *US$10 incentive* to participate. Applied in regular home environment though enumerators instructed to secure quiet space. Recorded data on administration conditions (time, duration, distraction, examiner FE)

Measuring Socio-emotional Traits: Big-Five Personality Factors

Big Five Factor	APA Dictionary description	NEO-PI-R facets (trait adjective)	Other related constructs
Conscientiousness	"the tendency to be organized, responsible, and hardworking"	Competence (efficient) Order (organized) Dutifulness (not careless) Achievement striving (ambitious) Self-discipline (not lazy) Deliberation (not impulsive)	Grit / Perseverance Delay of gratification Impulse control Self-efficacy
Neuroticism/ Emotional Stability	Neuroticism is "a chronic level of emotional instability and proneness to psychological distress." Emotional stability is "predictability and consistency in emotional reactions, with absence of rapid mood changes."	Anxiety (worrying) Hostility (irritable) Depression (not contented) Self-consciousness (shy) Impulsiveness (moody) Vulnerability to stress (not self-confident)	Self-esteem Internal locus of control Depression and related disorders
Agreeableness	"the tendency to act in a cooperative, unselfish manner"	Trust (forgiving) Straight-forwardness (not demanding) Altruism (warm) Compliance (not stubborn) Modesty (not show-off) Tender-mindedness (sympathetic)	
Openness to Experience	"the tendency to be open to new aesthetic, cultural, or intellectual experiences"	Fantasy (imaginative) Aesthetic (artistic) Feelings (excitable) Actions (wide interests) Ideas (curious) Values (unconventional)	
Extraversion	"an orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability"	Warmth (friendly) Gregariousness (sociable) Assertiveness (self-confident) Activity (energetic) Excitement seeking (adventurous) Positive emotions (enthusiastic)	

ANNEX B. Instrumental variables analysis

Performing traditional 2SLS for endogenous schooling

The more often used procedure to solve the omitted ability bias problem in the literature is to estimate the model using instrumental variables or 2SLS. The data used in almost all the labor literature on the returns to schooling do not include measures of ability. For this reason, we perform the estimation of the log earnings regression including schooling but not our measured skills using 2SLS to compare what we get from our data to results from previous studies.

The survey we use includes a module on school trajectories that permit us constructing several variables that affect acquired years of schooling and we can think these can serve as exclusion restrictions in the earnings equation. These potential instrumental variables are: scholastic achievement, effort exerted at school, time to get to school, assisted a preschool, parents placed a high value on education, and the mother placed a higher value on education than the father.

All these variables refer to the time the individual was pursuing her highest education level. The variables are constructed from self-reports to questions on the school trajectories module from the survey, and included as dummy variables that take a value of 1 for responses of high o highest response categories and a value of 0 in other cases. For instance, the scholastic achievement dummy takes a value of 1 when the individual obtained grades above the average, and a value of zero for grade below the average. An exception is the dummy for the mother placing a highest value on education. In this case the dummy takes a value of 1 when the mother and nor the father placed a high value to education.

Additionally, we also consider the socioeconomic status of the family at the time the individual was at school. The variable is constructed from the self-reported SES the individual declares her family had when she was in school. This variable is included as a dummy that takes a value of 1 when the SES was above the average and a value of 0 when SES was below the average. However, the assumption that the SES of the family at the time the individual was at school could affect earnings only through its effect on schooling is hardly convincing. Anyway, we include SES in the analysis for comparison purposes.

We implement this method in Table B. 1 using single instruments for schooling. Once we correct for the endogeneity of schooling, the point estimates we obtain from 2SLS for the return to schooling vary from 0 to 0.25 depending on the instrument we use. Only the estimates that use scholastic achievement and SES are statistically significant. The point estimates are 0.088 when the instrument is scholastic achievement and 0.25 when the instrument is SES. We also find that some of the instruments are weak. The instruments time to school, assisted to preschool, and mother placed a more value to education do not pass the test of weak instruments.

In addition, we also run the 2SLS regressions using grouped instruments, successively including additional instruments in the first stage regression. The results are reported in Table B. 2. All the estimates from 2SLS are statistically significant and vary from 0.069 to 0.094. Although the difference between the OLS and 2SLS are not large, these results are in line with what typical results in the labor literature: once schooling is instrumented, the estimated return to schooling by 2SLS is usually larger than the OLS estimate.

Schooling and cognitive skills endogenous

In the main text we assume that earnings depend on schooling and measured ability. An alternative formulation includes latent ability (A) in the true earnings model. Therefore, the true model would be written as:

$$\ln y = \alpha + \beta S + \gamma T + \phi A + \varepsilon. \tag{7}$$

Since we do not have access to measures of latent ability, this variable becomes an omitted variable in the estimated regression. Again, we can use instruments in the 2SLS method to account for the potential endogeneity of schooling and measured skills in the estimated regression model. The first-stage regressions are:

$$S = \pi_0 + \pi_1 X + \pi_2 Z + \nu, \tag{8}$$

and

$$T = \delta_0 + \delta_1 X + \delta_2 Z + \eta. \tag{9}$$

We have enough instruments in our data to accommodate the endogeneity of both schooling and cognitive skills, but not to accommodate for the endogeneity of non-cognitive skills. Thus, in the following analysis we include schooling and measured cognitive skills in the earnings regression but not measured non-cognitive skills.

Our main concern is that the instruments we have are related to schooling, but they are less likely independently related to post schooling measures of skills.

In Table B. 3 we report the results of our 2SLS regressions using grouped instruments for schooling and measured cognitive skills. Each of the columns from the second onwards includes one additional instrument in the first stage regression. The estimated coefficients for schooling are extremely large while the estimated coefficients for measured cognitive skills are all negative. With one only exception, none of these estimates is statistically significant. The only statistically significant coefficient is that for schooling in the last column. The magnitude of this coefficient is improbable. Even more, in the particular case of this last regression, the over identification J test suggest that some of the exclusion restrictions are not valid. This is the result of including SES as an additional instrument in the regression.

A problem with this 2SLS solution is that the model does not recognize that measured skills are affected by schooling. Even more, we can think of latent ability as a determinant of both schooling decisions and measured skills. A more complete model is written as:

$$S = \pi_0 + \pi_1 X + \pi_2 Z_S + \pi_3 A + \nu. \tag{10}$$

$$T = \delta_0 + \delta_1 X + \delta_2 S + \delta_3 Z_T + \delta_4 A + \eta. \tag{11}$$

$$\ln y = \alpha + \beta S + \gamma T + \phi A + \varepsilon. \tag{12}$$

The augmented model recognizes the dependence of schooling and measured skills on latent ability. It also recognizes the dependence of measured skills on acquired schooling. In this situation, however, we would need additional instruments that affect measured skills but not schooling in order to identify the model.

Table B. 1
Returns to schooling without controlling for skills
2SLS using single instruments for schooling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Method:	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Instrument for schooling:		Scholastic achievement	Effort exerted	Time to school	Assisted to Preschool	SES at highest level	Parents' valuation of education	Mother valued education most
Years of schooling	0.072***	0.088***	0.019	-0.283	-0.189	0.251***	-0.021	0.009
	[0.013]	[0.027]	[0.067]	[0.411]	[0.254]	[0.092]	[0.107]	[0.165]
Work exp.	0.323***	0.315***	0.349***	0.494**	0.449**	0.237*	0.368***	0.354***
	[0.104]	[0.106]	[0.112]	[0.251]	[0.184]	[0.142]	[0.128]	[0.137]
Work exp2 (x100)	-0.085**	-0.084**	-0.088**	-0.105*	-0.100**	-0.075*	-0.090**	-0.089**
	[0.036]	[0.037]	[0.037]	[0.057]	[0.048]	[0.044]	[0.039]	[0.039]
Sex	0.143**	0.135**	0.167***	0.307	0.264*	0.060	0.186**	0.172*
	[0.064]	[0.065]	[0.060]	[0.218]	[0.142]	[0.075]	[0.076]	[0.094]
Observations	1140	1140	1140	1140	1140	1140	1140	1140
R-squared	0.169	0.166	0.142	-1.044	-0.487	-0.139	0.087	0.130
Weak instruments		220.2	55.06	2.159	5.034	31.69	16.78	5.176
Under identification		97.87	30.48	0.954	2.945	20.65	9.477	2.618
Over identification		-	-	-	-	-	-	-

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table B. 2
Returns to schooling without controlling for skills
2SLS using grouped instruments for schooling

	(1)	(2)	(3)	(4)	(5)	(6)
Method:	OLS	2SLS	2SLS	2SLS	2SLS	2SLS
Instrument for schooling:		Achievement, effort	+ time to school	+ valuation of education	+ preschool	+ SES
Years of schooling	0.072***	0.077***	0.074***	0.073**	0.069**	0.094***
	[0.013]	[0.027]	[0.028]	[0.028]	[0.030]	[0.032]
Work exp.	0.323***	0.321***	0.322***	0.323***	0.325***	0.312***
	[0.104]	[0.106]	[0.106]	[0.107]	[0.107]	[0.109]
Work exp2 (x100)	-0.085**	-0.085**	-0.085**	-0.085**	-0.085**	-0.084**
	[0.036]	[0.037]	[0.037]	[0.037]	[0.037]	[0.037]
Sex	0.143**	0.140**	0.142**	0.142**	0.144**	0.132**
	[0.064]	[0.063]	[0.063]	[0.063]	[0.063]	[0.062]
Observations	1140	1140	1140	1140	1140	1140
R-squared	0.169	0.169	0.169	0.169	0.169	0.164
Weak instruments		123.5	82.94	50.52	42.55	41.30
Under identification		102.1	103.0	103.8	106.4	115.5
Over identification		1.050	3.178	3.292	3.705	14.41

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table B. 3
Returns to schooling and cognitive skills without controlling for non-cognitive skills
2SLS using grouped instruments for schooling and cognitive skills

	(1)	(2)	(3)	(4)	(5)	(6)
Method:	OLS	2SLS	2SLS	2SLS	2SLS	2SLS
Instrument for schooling:		Achievement, effort	+ time to school	+ valuation of education	+ preschool	+ SES
Years of schooling	0.054***	0.348	0.506	0.367	0.652	0.439*
	[0.014]	[0.358]	[0.633]	[0.483]	[2.019]	[0.255]
Aggregate cognitive measure	0.100***	-1.088	-1.730	-1.157	-2.299	-1.472
	[0.034]	[1.446]	[2.537]	[1.902]	[7.942]	[1.082]
Work exp.	0.311***	0.424*	0.485	0.430*	0.537	0.462*
	[0.103]	[0.228]	[0.347]	[0.246]	[0.778]	[0.240]
Work exp2 (x100)	-0.082**	-0.118	-0.138	-0.120	-0.155	-0.130*
	[0.036]	[0.073]	[0.109]	[0.080]	[0.253]	[0.074]
Sex	0.129**	0.254	0.322	0.261	0.380	0.296*
	[0.063]	[0.158]	[0.271]	[0.198]	[0.806]	[0.172]
Constant	0.134	-3.209	-5.009	-3.413	-6.642	-4.262
	[0.200]	[4.082]	[7.185]	[5.448]	[22.714]	[2.984]
Observations	1140	1140	1140	1140	1140	1140
R-squared	0.177	-1.042	-2.715	-1.190	-4.805	-1.951
Weak instruments		1.137	0.772	1.165	1.009	1.987
Under identification		1.522	1.525	4.115	4.216	7.928
Over identification		-	0.300	1.115	0.738	2.219

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1