Returns to cognitive skills in 7 developing countries

Marcello Perez-Alvarez
University of Göttingen, Germany
mperez@uni-goettingen.de

December 2017

Preliminary Draft

Abstract

Empirical support for the human capital theory across the developing world predominantly uses years of schooling as the indicator for human capital. However, the learning crisis taking place in developing countries strongly suggests that schooling quantity represents an inferior proxy for this type of analysis (World Bank, 2018). This paper relies on cognitive skills proxied by literacy skills to estimate the returns to human capital for wage workers in the urban areas of 7 developing countries. By implicitly taking into consideration school quality aspects and non-school influences in human capital formation, cognitive skills better represent the rationale of human capital theory. The results in the cross-country specification show that, on average, an increase in one standard deviation of literacy scores increases net hourly wages by 8.5%, conditioning on years of schooling. The findings are robust to the inclusion of covariates related to socioeconomic background and alternative human capital indicators. Interestingly, returns to skills vary greatly across developing countries. This analysis implies that the empirical support for the human capital theory is remarkably heterogeneous across the developing world.

Returns to cognitive skills in 8 developing countries

1. Introduction

Investments in human capital have been widely recognized as a major determinant of earnings. Based on human capital theory, such investments increase individuals' earnings by raising their working productivity (Schultz, 1961; Becker, 1962; Mincer, 1970, 1974). Empirical support for this statement has overwhelmingly relied on schooling quantity as the central indicator for human capital (Card, 1999; Patrinos and Psacharopoulos, 2004; Montenegro and Patrinos, 2014). Recently however, there has been a shift towards the use of cognitive skills measures instead (see Glewwe (2002) and Hanushek and Woessmann (2008) for a review and Hanushek et al. (2015) for estimations across developed countries).

This paper uses the World Bank STEP Skills Measurement adult skills surveys to estimates the returns to cognitive skills for wage workers in the urban areas of the countries of Armenia, Bolivia, Colombia, Georgia, Ghana, Kenya, Ukraine and Vietnam. The proxy for cognitive skills are literacy skills, defined as the capacity of "understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential" (OECD, 2013, p. 59). While Valerio et al. (2016) perform a similar exercise, this analysis deviates in four aspects. First, in order to compare the estimations with those of Hanushek et al. (2015) for developed countries, I restrict the sample to full-time wage workers, leaving the self-employed out of the analysis. Second, I systematically test the robustness of the results against the inclusion of covariates related to socioeconomic background and alternative human capital indicators. Third, I employ a 2SLS-IV estimation as an additional robustness check. Finally, I investigate transmission channels.

Unlike schooling quantity, the use of cognitive skills measures allows for differentiating between competing theories of the determinants of earnings. While signaling (Spence, 1973, 2002), credentialist (Bills, 2003) and Marxist (Bowles and Gintis, 1975) approaches also predict a positive schooling-earnings relationship, they do not necessarily imply a positive relationship between cognitive skills and earnings (Boissiere et al., 1985, p. 1016; Alderman et al., 1996, p. 34; Glewwe, 2002, p. 466).

Moreover, cognitive skills measures implicitly contain information on the fundamental role of schooling quality and non-school influences in the production of human capital, which were previously neglected by standard estimations of returns to schooling (Hanushek and Woessmann, 2008, pp. 611-612). Schooling quantity expansions alone are no guarantee for individuals' productivity increases and hence imperfectly represent the rationale of the human capital theory, particularly in economic analyses covering developing countries. Recently, in a report covering 85 countries, it was estimated that 160 million children with at least 4 years of primary schooling were not able to read and comprehend a simple sentence, a finding catalogued as "a global learning crisis" (World Bank, 2018; UNESCO, 2014, p. 191). This highlights the relevance of school quality aspects for earnings analyses covering developing countries.

Results of this analysis show a positive association between cognitive skills and wages. In the pooled estimation that controls for years of schooling, a one standard deviation increase in literacy skills is on average related to a 8.5% net hourly wage increase. Exceptions to this general trend are the Kenyan and Vietnamese regressions, where coefficients on literacy scores are small and insignificant. This suggests that competing theories of the human capital framework might find empirical support in specific labor markets. Returns to literacy estimated via the 2SLS-IV technique increase to 44.5%, suggesting that OLS estimations are more affected by attenuation bias rather than by omitted variable bias.

The structure of this paper is as follows. Section 2 reviews the related literature. Section 3 starts by presenting the data, sample delimitation and variables description, followed by descriptive statistics of the key variables. Section 4 presents the empirical framework and the results. Special attention is given to omitted variable bias concerns and the economic reasoning behind the choice of control variables. A 2SLS-IV estimation is carried out as an additional robustness check. In addition, transmission channels are investigated. Section 5 concludes.

2. Related literature

2.1. Years of schooling and cognitive skills

Most of the literature assessing individual earnings portrays them as a positive function of human capital[1] (Card, 1999; Borjas, 2013, pp. 235-282). Within this approach, individuals undergo investments in human capital expecting higher earnings, resulting from increases in worker's productivity (Schultz, 1961; Becker, 1962; Mincer, 1970, 1974).

Mincer (1970, 1974) provides the theoretical foundation for the decomposition of human capital into schooling and on-the-job training, with focus on the former. This insight gave birth to the canonical human capital earnings function, which estimates the impact of years of schooling on earnings and

[1] The terms *human capital* and *labor market skills* are used interchangeably throughout this study.

prevails among the most widely used empirical models in the economic literature (Card, 1999; Lemieux, 2003).

Hanushek and Woessmann (2008, pp. 609-615) review and motivate a more recent framework of the human capital approach, which replaces schooling with cognitive skills as the central indicator of labor market skills. Cognitive skills are measured via standardized evaluations of literacy, numeracy, science and problem-solving capacities. Although the theoretical background of the human capital theory remains unchanged, the empirical implications and interpretations clearly diverge from the Mincerian approach. Within this framework, schooling affects earnings solely through cognitive skills.

The comparative advantages of the cognitive skills approach are notable: first, cognitive skills measures implicitly contain information on the fundamental role of schooling quality and non-school influences such as innate ability and health and family inputs in the production of human capital, previously neglected by standard estimations of returns to schooling. A corresponding comprehensive data collection of these input factors would be an impractical if not an impossible alternative. In this view, schooling quantity is just one of many factors contributing to the acquisition of cognitive skills. Notably, that perspective is consistent with the large literature on education production functions (Hanushek and Woessmann, 2008, pp. 611-613).

School quality is intrinsically connected to the rationale of the human capital theory. Unfortunately, the standard Mincerian wage equation does not consider that dimension of schooling, and by doing so implicitly assumes that a specific year of schooling, irrespective of its quality, increases the individuals' productivity by the same amount (Hanushek and Woessmann, 2008, pp. 608). That implication is strongly contradicted by the empirical evidence across countries. For instance, while the completion of primary schooling in developed countries ensures that children are able to read and write, this is not necessarily the case in developing countries. It has recently been estimated that 20 million children in developing countries with 6 years of schooling completed are not able to read a simple sentence (UNESCO, 2014, pp. 208-209). In the same spirit, international student assessments such as PISA, PIRLS and TIMMS point out the great variation in achievement within and between countries among students with the same years of schooling (Leuven et al., 2004, pp. 5-6).

A second and related benefit of this approach is its focus on an output variable rather than an input variable (Alderman et al., 1996, p. 34). This feature is advantageous because it takes into account the heterogeneity that individuals might show in their capability to transform the same schooling input into different levels of cognitive skills. Finally, competing theories of the role of education - human capital, signaling, credentialism - share empirical implications when using schooling quantity as the central indicator for human capital, making it hard to test their validity separately. On the other hand, when using cognitive skills measures, the empirical predictions of the human capital theory do not coincide with the competing frameworks (Boissiere et al., 1985, p. 1016; Alderman et al., 1996, p. 34; Glewwe, 2002, p. 466). That is relevant as the above-mentioned theories differ in their policy implications.

2.2. Previous empirical analysis

Estimated returns to cognitive skills in developing countries are too few to establish general patterns, and in particular, whether cognitive skills returns show decreasing marginal returns by stage of development of countries. Behrman et al. (2009) uses a longitudinal dataset of rural Guatemala to estimate returns to cognitive human capital, proxied by reading comprehension test scores, and health human capital, proxied by height and body mass index. They estimate returns to reading comprehension of 22.8%. Additionally, a 2SLS IV-GMM estimation was carried out. Estimated returns to reading comprehension amount to 40.4%.

Boissiere et al. (1985) estimate returns to a combined score of literacy and numeracy tests in urban Kenya and urban Tanzania controlling for a schooling dummy for primary and secondary completers. The separate use of either of the scores does not change their results. The returns vary from 19% to 22% for Kenya and from 7% to 13% for Tanzania, depending on selected subsamples. Jolliffe (1998) estimates the association of household income and mathematics and English tests scores in Ghana with the national Ghana Living Standards Survey. Combined returns amount to a statistically significant value of 9.6% for total income, without controlling for schooling. Estimates controlling for schooling are unfortunately not performed.

Glewwe (1996) uses a subsection of 359 observations from the same dataset used by Jolliffe (1998). He focuses on wage earners and corrects for selectivity bias, a potential problem for generalizations of wage equations results from Ghana, given the high incidence of self-employment in the country. The Heckman correction for selectivity however does not change the results. Estimated returns vary from a lower bound of 14% for private wage workers to an upper bound of 30% for government wage workers, without controlling for years of schooling. After controlling for years of schooling, returns to mathematics and English scores are reduced by roughly 14%. Returns to cognitive skills measures in the private sector are statistically significant only at the 10%, while in the public sector they are significant at the 1% level. The coefficient on years of schooling is positive and significant only for wage workers in the public sector. The author argues that this finding is evidence for wages in the government sector being more credentialist than in the private sector.

Alderman et al. (1996) uses data from rural Pakistan to estimate returns to a combined score of literacy and numeracy between 12% and 28%, with the latter being statistically significant. These results control for years of schooling and height and are robust to the separate use of each of the scores. The analysis deals with the endogeneity of the cognitive skills measures using Maximum Likelihood IV estimation. Scores are instrumented with a set of predetermined variables presumed to influence investments in human capital during childhood, such as distance to school, textbook costs, village dummies and parental schooling.

Foreign language skills have also been used as a proxy for human capital. Angrist and Lavy (1997) regress wages on schooling, French skills test scores and other standard regressors of wage equations. They estimate a negative impact of reduced French language skills in Morocco and a positive effect of schooling attendance on wages. They make use of an institutional change in the schooling system of Morocco, which was the switch in the language of instruction from French to Arabic in 1983, to compute 2SLS IV estimates, which are larger than OLS estimates. Both OLS and IV estimates on schooling and test scores are positive and statistically significant.

An instructive cross-country analysis presented by Hanushek et al. (2015) present higher estimates than the above-mentioned values. The assessment makes use of the cross-section database from the Programme for International Assessment of Adult Competencies (PIAAC) to estimate returns to cognitive skills across 22 OECD countries. In the pooled regression, the returns to numeracy skills take the value of 18% for full-time wage workers aged 35-54. The range of estimated national returns covers a wide range of 12%-28%, which are always highly statistically significant. When using literacy skills as the central measure for cognitive skills, returns are very similar. Three additional control variables are separately used to ease omitted variable concerns. Among them, years of schooling, which always enters the equation with positive and highly significant coefficient, lowers the coefficient on numeracy the most, to a value of 10% in the pooled regression. This drop is interpreted by the authors as a suggestion that cognitive skills affect earnings through expansions in schooling. The empirical work of this study is rooted in this framework. It emulates its criteria for sample selection, expands the set of control variables and adapts it to the developing context.

There are several conclusions to be drawn from these studies. First, cognitive skills measures seem to be positively associated with earnings in both developing and developed countries. Second, results are robust to the choice of the cognitive skills dimension, numeracy or literacy. Third, the coefficient on years of schooling, conditioned on cognitive skills measures, is mostly positive. Boissiere et al. (1985) and Glewwe (1996) attribute this to the credentialist role of schooling, while Heckman et al. (2006b) favor the view that noncognitive skills are the driving force behind it. Finally, Hanushek et al. (2015) acknowledge that the interpretation is undetermined, but subtlety favor the hypothesis that measurement error in cognitive skills lies behind this finding. Most likely, all three interpretations might be playing a role. Fourth, it seems that whenever a proxy for innate ability - Raven's test score - is added to a wage equation that include a cognitive skills measure - literacy or numeracy test scores -, the former does not have an impact on earnings. All three studies interpreting this result conclude that the screening theory is relegated to the benefit of the human capital theory.

Finally, all these estimates support the cognitive skills approach of the human capital theory, especially after considering that test scores might be plagued by standard measurement error generating attenuation bias. Yet, comprehensive identification strategies dissipating omitted variable bias concerns have not been carried out.

Recently, it has been increasingly argued that noncognitive skills in the form of personality traits play an important role in generating cognitive skills and earnings, and as a result might represent omitted variables in the estimations of the studies mentioned above (see Bowles et al. (2001), Cunha et al. (2006, pp. 717-719), Brunello and Schlotter (2011), Heckman et al. (2006b) and Lavado et al. (2013)). Therefore, controlling for these factors is necessary to further ease endogeneity concerns.

3. The STEP Data

This study uses the World Bank's Skills Towards Employability and Productivity program (STEP). The dataset consists of representative household surveys carried out in 7 developing countries: Armenia, Bolivia, Colombia, Georgia, Ghana, Kenya and Vietnam.[2] The implementation of the surveys took place in the years 2012 and 2013. The population surveyed consists of individuals living in urban areas aged 15-64 (World Bank, 2014, pp. 7-9). Individuals were randomly selected with a multistage sampling procedure.

This rich cross-sectional dataset is the first of its kind. It is the first in combining internationally comparable cognitive skills measures of the working-age population in various developing countries with variables related to individual labor market outcomes. Individual information on educational attainment, personality traits, health, language skills, family and socioeconomic background, among others, are also collected. An additional profitable feature of this database is its comparability to the Programme for International Assessment of Adult Competencies (PIAAC), carried out in developed

[2] The program also implemented household surveys in Ukraine, Lao PDR, Macedonia, Sri Lanka and the Yunnan Province in China. Unfortunately, the skills assessments in the latter 4 were not extensive enough for proper regression analysis. (World Bank, 2014, p. 8, 38).

countries. The STEP's surveys were purposefully based on the PIAAC database to enable proper comparison (World Bank, 2014, p. 13, 27, 35).

I reduce the database to a subsample of full-time wage workers aged 30-59. In accordance with the International Labor Organization (ILO), I define full-time wage workers as those employees working more than 30 hours per week, or less than 30 hours per week, but unwilling to work more hours at their current hourly wage (ILO, 2013b, p. 41, 55; Lee et al., 2007, pp. 58-60). In the overall pooled dataset, 21,242 observations are available. From those, 14,928 individuals are active in the labor market. In reality, only the 12,500 employed individuals among them represent the eligible sample. Notice that labor market inactivity is a much larger phenomenon than unemployment. Given that profits from self-employment are the outcome of a fundamentally different process with respect to wages, only the 7,039 observations corresponding to wage workers are taken into consideration. I further restrict the sample to ages in the range of 30-59, which makes the number of observations drop to 4,151. The reason of this restriction is that individuals' earnings within this age range have proven to better proxy lifetime earnings. In addition, the human capital-wages relation at younger ages might be distorted by skills mismatch and the fact that skills need time to become observable and therefore be rewarded by the employer (Hanushek et al., 2015, p. 7).

Among the 4,151 wage workers aged 30-59, those working full-time amount to 3,920 individuals. It is important to note that this value includes formal and informal wage workers. Ultimately, the number of observations for the baseline regression is slightly reduced to 3,728 observations by the exclusion of observations with typing errors, uncommon missing values, and the trim of the 1st and 99th percentile observations of the wage distribution, as well as by the drop of one outlier in the cognitive skills-earnings relation in the Ghanaian dataset. In sum, the major reductions in the number of observations come firstly from taking self-employed individuals out of the sample, and secondly from the age range restriction. In particular, the exclusion of the self-employed could potentially be causing a selection bias. However, by implementing this restriction, a more homogenous group of individuals is selected, and the results of the analysis are more comparable to similar studies for developed countries.

It is important to acknowledge that the previous delimitation of the sample might raise issues of selection bias (Heckman, 1979, pp. 153-154). A natural byproduct of the analysis of wages is the exclusion of observations corresponding to the unemployed and the self-employed. Excluding unemployed individuals from the sample is associated with a downward bias on returns to human capital measures, as human capital is presumed to positively affect the likelihood of being employed (Puhani, 2000, p. 53). Hence, in this respect estimations of returns to cognitive skills can be considered as conservative. It is however less clear how the exclusion of the self-employed might affect estimations. While the choice of self-employment vs. wage employment is surely nonrandom (Glewwe, 2002, p. 467; Rosenzweig, 2010, p. 4), the direction of the selection bias is unclear. Still, if cognitive skills are positively related to wage employment choice, the selection bias is unlikely to be upward (Alderman et al., 1996, p. 38; Glewwe, 2002, p. 467). Moreover, previous studies on returns to human capital measures in developing countries do not find significant coefficient movements after attempts to correct for this selection bias (Alderman et al., 1996; Glewwe, 1996; Soon, 1987).[3]

I now turn to the description of the main variables to be used in the regression, starting with the work-related ones. The variable net hourly wage was generated by dividing the total wage income by the total hours of the main occupation, both values corresponding to the last month/week previous to the interview. Wage information from second and third jobs is not available. This is not a major problem, as 93.1% of the selected sample has a single job.

The hourly wages are net of taxes and social insurance deductions. As a result, differences in taxation rates might partly influence differences in the earnings of formal wage workers. Hourly wages are available in national currencies. For the cross-country analysis, they have to be cautiously converted to a common monetary unit with the same base year. For that purpose I use the PPP conversion factors for private consumption from the World Development Indicator (WDI) series available at the World Bank database. For surveys carried out towards the end of 2012, the conversion is achieved by the use of the PPP conversion factors of 2012. For surveys implemented towards the end of 2013, a deflator I constructed with the household final consumption expenditure series of the WDI has been used to deflate wages from 2013 into wages of 2012 in local currency. In the second step, the deflated wages were converted into international dollars using PPP conversion factors of 2012 (World Bank, 2015, p. 172; Schreyer and Koechlin, 2002, pp. 5-7).

Other useful work-related variables are occupation, classified into 9 types according to the International Standard Classification of Occupations (ISCO) 2008, economic sector of the job, classified into 21 sectors according to the International Standard Industrial Classification of All Economic Activities (ISIC) 2008, and whether the worker has a formal job or not, defined as being affiliated with social security benefits at work. Additionally, information on whether the worker is employed in the public or private sector and employment search methods is collected.

[3] A third potential type of selection bias arises from the heterogeneous response rates of the surveys. In Bolivia and Colombia, only 43% and 48% of randomly selected individuals completed the survey, while 83% and 92% did so in Ghana and Kenya (World Bank, 2014, pp. 61-62).

Turning to the main independent variable of the analysis, World Bank (2014), the designer of the STEP dataset, define cognitive skills as “the ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought”[4] (Neisser et al., 1996, p. 77). The chosen proxy by STEP designers for the latent cognitive skills is literacy, defined as “understanding, evaluating, using and engaging with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential” (OECD, 2013, p. 59). Despite the generality of both concepts, one can identify major overlaps between the two.

The reading literacy scores are derived from an assessment that contains exercises emulating the diversity of real-world tasks in an adult’s daily life, bringing together aspects of work, education, health, home, and citizenship, among others. It contains items from the PIAAC and is adapted to the context of developing countries. It is noteworthy that for international comparison purposes items are designed to be culture and language neutral.

Anticipating a wide variation in literacy skills in developing countries, the test design provides different levels of difficulty, ranging from very easy to very difficult items. In the first stage, individuals face basic tasks aiming at measuring foundational skills. Subsequently, a sorting stage takes place. If the respondent answers 3 out of 8 questions correctly, she proceeds to the third and final stage of the test, which covers the full range of difficulty. For respondents failing at the screening stage, incorrect responses for the third stage are imputed. Based on previous pilot tests, this is an accurate imputation. The first stage of basic difficulty aims at identifying variation at the left tail of the skills distribution, while the third evaluates the full variation for those equipped with more than very basic literacy skills. The overall test duration amounts to 45 minutes (ETS, 2014, pp. 5-7, 24).

The STEP literacy scale is fully comparable to the PIAAC scale, with scores denoted as plausible values, ranging from 0 to 500 points (ETS, 2014, pp. 11-12). Two major techniques were implemented to generate the plausible values, which are multiple imputations. Given that different groups of individuals solved different task items, it is not appropriate to base the scores solely on the number of correct and incorrect responses. The Item Response Theory (IRT) scaling is used to weigh each item’s contribution to the final score based on their level of difficulty and to impute responses to

[4] Neisser et al. (1996) use this concept to describe intelligence in the psychology literature. World Bank (2014) and I apply this concept to define cognitive skills in the economic literature. Somewhat odd, cognitive skills, unlike cognitive skills proxies, are seldom defined in the reviewed labor economic literature. In a nutshell, the neuroscience literature understands as cognitive skills the ability to represent, store, and process information in the cognitive domain (Gazzaniga, 2009, pp. 805-806). OECD (2013, pp. 56-59) implicitly defines cognitive skills as key information-processing skills, related to accessing, identifying, evaluating, using and communicating information.

items that were not administered to a certain individual. These imputations are based on the individuals' profile, derived from the set of answered items. In this way a score distribution is generated. The second issue is the tendency of large-scale cognitive assessments to administer a relatively low number of items in order to reduce the respondent's burden. As a result, an otherwise avoidable measurement error type arises. A latent regression model combining IRT values and individual background variables such as country of birth, education and occupation is used to minimize this measurement error. This technique is implemented to enhance the IRT skills distribution, from which 10 plausible values per respondent are derived. For the correct use of the plausible values in regression analysis, it is suggested to run 10 separate regressions with each plausible value. The mean coefficient of the 10 estimations is the consistent coefficient of interest (World Bank, 2014, pp. 95-98; ETS, 2014, pp. 26-31; Junker et al., 2012, pp. 6-8). It is important to mention that these procedure attempts to correct for measurement error related to test-retest reliability. A more fundamental measurement error inherent to the broad concept of literacy, cognitive skills and ultimately human capital, remains (Hanushek et al., 2015, p. 6).

The International Standard Classification of Education (ISCED) 1997 and 2011 scales are used to derive individual years of schooling. This variable represents the number of years of formal education corresponding to the highest level completed. In Bolivia, Colombia, Georgia and Vietnam, intervals of levels completed are narrow, such that one additional year often coincides with one level increase. In Armenia, Ghana and Kenya, however, the intervals of established levels in the education system are large, such that relevant years' variation gets lost. Hence, for these three countries, the actual years of formal education completed are taken. This variable captures primary, secondary, vocational and university education.

Special attention to the context of developing countries has to be taken into consideration to derive a variable for potential labor experience. For developed countries analysis, experience is calculated by computing age minus years of schooling minus 6, assumed to be school starting age. In doing so, the full time period after schooling is attributed to labor experience. Adopting this definition for developing countries is problematic because a non-negligible share of individuals attends formal schooling for a few or 0 years and start working at very young ages. Specifically this work at very young ages cannot be considered as human capital in the form of job experience. For instance, a 40 year old worker with no formal schooling cannot be considered to have accumulated job experience for 34 years. In addition, school starting age in developing countries presents important variation. Taking these two aspects into consideration, I compute experience with the operator min{age-years of schooling-actual school starting age; age-13}. That way, working years before age 13 are never considered as contributing to human capital (Alderman et al., 1996, p. 36; Lam and Levison, 1992, pp. 238-239).

The rich STEP dataset contains valuable variables that can be used as controls in wage regressions. To derive mother's education, the International Standard Classification of Education (ISCED) 1997 scale is used for comparison purposes. This is a categorical variable with 7 ascending levels of education. This alternative was preferred over actual years of parental schooling to avoid measurement error, as the accuracy of the latter is questionable given that parental education goes several years back in time and refers to a different person than the respondent. Self-reported socioeconomic status of individuals at age 15 on a scale from 0 to 10, with 10 being the wealthiest level, is also collected, as well as the height of individuals. Other valuable predetermined variables used in the empirical strategy, such as kindergarten attendance, self-reported academic performance, and economic shocks at age 12 are detailed in the appendix.

Two further variables worth commenting are language proficiency and self-reported noncognitive skills. The former refers to the number of languages that individuals speak well enough to work in a job requiring that language. The latter are a set of 5 well-known measures in the psychology literature denominated the "big five personality traits": Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. Each trait measure is self-reported by answering a set of 3 questions designed to uncover the traits. For example, to measure emotional stability, the individual answers to the question "*do you get nervous easily?",* on a scale from 1 - almost never - to 4 - almost always - (World Bank, 2014, pp. 28, 77-78).

Table 1 shows the mean and standard deviation of the most relevant variables for the restricted sample. The first column shows the descriptive statistics for the pooled sample, followed by columns of country samples. Mean hourly wages are shown in international dollars and national currency, with only the former being comparable. The average hourly wage takes a low value of 3.49 international dollars. This value amounts to less than half the nationwide minimum wage in the United States (OECD, 2014). There is important variation between countries. Its distribution, as commonly reported in previous analysis, is skewed to the right in all samples (Berndt, 1996, p. 161). Ghana presents the lowest hourly wage, while Bolivia shows the highest. Bolivia also presents the highest variation around the mean, suggesting large labor income disparities among urban wage workers. It draws attention that Armenia has the second lowest mean. There can be many explanations for this. Perhaps the wage sector in Armenia faces a costly bundle of consumptions goods. Another explanation could be that tax and social security payments are relatively larger for the Armenian wage sector, given that informality plays a minor role in its urban wage sector.

The informal wage sector is as expected relatively large. Every third urban employee in the sample works in the informal sector. Variation across countries is tremendous. Armenia shows an almost negligible 6.3%, while more than half of Georgian and Ghanaian workers in the sample are informal employees. The same disparities dimension is observed in the share of public workers. 70% of Armenian wage workers in the sample function in the public sector, and only 12% do so in the Colombian dataset.

In general, mean literacy scores are low. The mean of 225 points is several tenths below the mean literacy score of 273 points from the PIAAC study covering adults in 23 developed countries (OECD, 2013, p. 258).[5] The distribution of the scores is mostly approximately normal, with a thin left-hand-side tail representing those individuals with very basic literacy skills sorted out at the screening stage of the test. For the cases of Bolivia, Ghana and Kenya, this latter group of individuals is numerous enough to transform the distribution shape into a bimodal distribution.[6] Again, a wide variation between countries is observed, with scores in Ghana almost 100 points lower than in Armenia. Vietnam on the other hand, performs two tenths better than the average.

Although differences in literacy scores between the PIAAC and the restricted STEP sample are large, mean years of schooling differ by less than 1 year (Hanushek et al., 2015, Table 1). That raises the relevance of differences in school quality and non-school investments in literacy skills. Another distinctive feature of the dataset is the low percentage of females among wage workers in Ghana and Kenya, in contrast with Armenia and Georgia. Age and experience are similar across samples. Father's education levels are not surprisingly higher than mother's education levels in all samples. However, both mean values of the pooled sample correspond to the same category of lower secondary education.

[5] One could argue that this comparison is not fully neat as the sample restrictions imposed to the STEP dataset might be lowering the mean score. However, the mean literacy score of the complete STEP dataset is even lower than the value shown in the table, amounting to 211 points.
[6] See Figure 1 in Appendix.

Table 1. Descriptive statistics

	Pooled	Armenia	Bolivia	Colombia	Georgia	Ghana	Kenya	Vietnam
Net hourly wage (international $)	3.5 (3.3)	2.8 (1.8)	5.2 (4.7)	4.1 (4.2)	3.8 (3.1)	2.2 (2.2)	3.3 (3.7)	3.3 (2.5)
Net hourly wage (local currency)	-	545.5 (353.8)	15.6 (14.1)	4901.5 (5046.0)	3.1 (2.5)	2.7 (2.8)	131.0 (149.1)	26900.4 (20606.5)
Share informal workers (%)	34.3 (0.5)	6.3 (0.2)	42.4 (0.5)	26.0 (0.4)	52.3 (0.5)	39.2 (0.5)	51.9 (0.5)	30.5 (0.5)
Share public workers (%)	41.6 (0.5)	70.2 (0.5)	28.6 (0.5)	12.1 (0.3)	58.1 (0.5)	39.5 (0.5)	15.6 (0.4)	53.6 (0.5)
Literacy score	224.9 (71.9)	257.7 (30.0)	203.5 (79.6)	234.6 (56.4)	247.6 (39.8)	166.5 (93.6)	179.7 (84.4)	247.0 (59.2)
Years of schooling	12.4 (4.4)	13.4 (2.6)	12.9 (4.8)	10.6 (4.1)	15.8 (2.7)	10.9 (4.5)	10.8 (4.5)	12.1 (4.4)
Experience	22.2 (9.1)	24.3 (9.5)	21.1 (9.0)	23.3 (9.1)	21.4 (8.7)	21.7 (9.3)	20.2 (8.5)	22.6 (9.1)
Age	41.4 (8.4)	44.6 (8.8)	40.5 (8.0)	41.1 (8.1)	43.6 (8.6)	39.9 (8.5)	38.4 (7.5)	41.4 (8.1)
Share of female workers (%)	51.0 (0.5)	65.5 (0.5)	49.2 (0.5)	49.8 (0.5)	66.1 (0.5)	31.8 (0.5)	33.3 (0.5)	53.6 (0.5)
Father's education (ISCED scale)	2.5 (1.9)	3.5 (1.3)	2.2 (1.9)	1.8 (1.4)	4.5 (1.5)	1.9 (1.8)	1.1 (1.6)	2.2 (1.7)
Mother's education (ISCED scale)	2.0 (1.9)	3.5 (1.2)	1.5 (1.8)	1.3 (1.0)	4.3 (1.5)	1.0 (1.4)	0.7 (1.4)	1.6 (1.5)
N	3727	568	413	496	492	351	541	866

Notes: Average values and its standard deviations in parenthesis. Bottom row shows number of observations. Sample selection: Full time urban wage workers aged 30-59. 1st and 99th percentile of hourly wage distribution trimmed. Net hourly wage in international dollars were calculated using PPP conversion factors for private consumption. Informal workers are defined as those who are not affiliated with social security benefits at work. Literacy scores are the mean of 10 plausible values, on a scale of 0 to 500. Experience is calculated with the following operator: min{age-years of schooling-actual school starting age; age-13}. Father's and mother's education levels are scaled according to the International Standard Classification of Education (ISCED) 1997 scale. Source: STEP database.

4. Empirical framework

The empirical framework I employ emulates part of the analysis in Hanushek et al. (2015), who undertake a similar approach for developed countries. The here employed STEP dataset shares plenty of similarities with their selected PIAAC database. In particular, the cognitive skills measures of both datasets are purposefully based on common items, methodology and scoring (World Bank, 2014, p. 61). I attempt to exploit these similarities and follow the above-mentioned framework closely to enable consistent comparisons between estimations for developed and developing countries.

Meaningful adaptations to the developing context and useful extensions enabled by the richness of the dataset are carried out. The following subsections start delimiting the estimation approach, which is directly followed by a presentation and description of the results.

4.1. Baseline estimation

I first run the baseline regression depicted in (1), with log earnings (ln Y) as a function of literacy skills (L), potential experience (E), potential experience squared (E^2), gender (F), geographical location (G) and a stochastic error (ε):

$$\ln Y = \beta_0 + \beta_1 L + \beta_2 E + \beta_3 E^2 + \beta_4 F + \beta_5 G + \varepsilon \qquad (1)$$

This equation is very similar to the standard human capital earnings function, with the difference that the proxy for human capital is literacy skills instead of years of schooling, for the reasons commented in section 2. The private returns to literacy[7] (β_1), also understood as the private returns to cognitive skills, are the focus of this empirical work. For comparison purposes, as is common in the literature I standardize the literacy scores to have mean 0 and standard deviation 1. Therefore, marginal returns correspond to a one standard deviation increase in scores. To put things in perspective, it is useful to understand what a standard deviation in literacy scores represents. For example, in the Colombian dataset, a one standard deviation from the mean in the literacy skills implies to jump one level higher in a scale of 6 levels. That implies that the individual passes from being able to employ low-level inferences including comparison, integration and identification from different parts of the test, to being able to identify and formulate responses in multisteps and differentiate correct from incorrect information (World Bank, 2014, p. 83).

The functional form of equation (1) depicts a log-linear wages-literacy skills relationship. That functional relation between wages and human capital measures is advantageous because it is based on a theoretical background.[8] Moreover, the logarithmization of hourly wages transforms its positively skewed distribution towards a normal distribution (Berndt, 1996, p. 161), and reduces heteroskedasticity concerns (Verbeek, 2012, pp. 84-85). In addition, the functional form favors comparability of coefficients, as they directly provide percentage changes in hourly wages for small changes in literacy skills units (Stock and Watson, 2007, p. 271).

[7] The concept returns to literacy corresponds to the average growth rate of earnings associated with a one unit increase of literacy skills, and not to internal rates of return.

[8] The core of the derivation relies on the assumption that individuals undergo investments in education only until they pay off in terms of rates of return, creating in that way a relation between linear human capital measures and percentage increases of earnings (Mincer, 1970).

The remaining three independent variables are commonly used in earnings equations. Potential experience enters the equation once linearly and once quadratically due to its presumed concave relationship with log earnings (Card, 1999, p. 1804). The gender dummy accounts for the well-known gender-based effects on wages and geographical location takes into consideration region-level fixed effects.

Table 2 reports the results from estimating equation (1). The empirical model is estimated for the pooled sample and for each country sample. In the country regressions, net log hourly wages in local currencies are regressed on literacy scores, which are standardized to have mean 0 and standard deviation 1 within each country. In the pooled regression, net log hourly wages in international dollars are regressed on literacy scores, again standardize to have mean 0 and standard deviation 1 but within the pooled international sample. All regressions include controls for experience, experience squared, gender and geographical location. Standard errors are clustered at the municipality level for Colombia and at the level of census enumeration areas for the other countries, to permit for intragroup correlation of errors.

Table 2. Baseline regression: Returns to Literacy

	(1) Pooled	(2) Armenia	(3) Bolivia	(4) Colombia	(5) Georgia	(6) Ghana	(7) Kenya	(8) Vietnam
Literacy	0.285***	0.083***	0.317***	0.262***	0.171***	0.386***	0.271***	0.190***
	(0.018)	(0.029)	(0.043)	(0.035)	(0.034)	(0.053)	(0.048)	(0.024)
N	3727	568	413	496	492	351	541	866
R^2	0.240	0.194	0.217	0.253	0.183	0.241	0.173	0.165

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for a constant, gender, experience and experience squared. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

The results in Table 2 consist of positive, large and highly significant returns to literacy for all samples. In the pooled regression, a one standard increase in literacy scores is associated with a 28.5% increase in net hourly wages on average. The standard deviation of the returns to literacy is lower in the pooled sample than in the country regressions. Interestingly, there is a wide variation of returns between countries. The Ghanaian and Bolivian samples show the highest returns to literacy, with above-average values of 38.6% and 31.7% respectively. Colombia and Kenya, with returns to literacy of 26%-27% show similar magnitudes to the pooled sample estimation. Georgia and Vietnam have returns between 17%-19%, while the Armenian sample presents the lowest return of 8.3%, which is nevertheless considerably high. The standard deviations of Bolivia, Ghana and Kenya are slightly above 4 percentage points, while the other coefficients are more precisely estimated. These estimations are prone to attenuation bias due to measurement error in literacy skills and hence should at first be taken as a lower bound estimate. In the pooled sample, the empirical model is able to explain 24% of the log hourly wage variation, whereas in country samples the R-squared values are slightly lower.

These results arguably show a tendency for falling returns by levels of development, as reported in cross-country analysis for returns to schooling (Psacharopoulos, 2004, pp. 112-113). In general, the reported coefficients in Table 2 are larger than the equivalent empirical estimations for developed countries in Hanushek et al. (2015), who report returns to literacy skills of 17.1% for the pooled sample.

4.2. Robustness checks

The major threat to the causal interpretation of returns to literacy reported in Table 2 concerns standard omitted variable bias.[9] Variables positively correlated with literacy skills might at the same time affect hourly wages positively.[10] Omitting these variables from the wage equation biases the returns to literacy upwards. In that case, returns to literacy should be interpreted as the combined effect of literacy and omitted variables on hourly wages. In the worst scenario, this upward bias could generate a spurious relationship between literacy and wages (Stock and Watson, 2007, pp. 187-191).

The main robustness check I employ to tackle these omitted variable bias concerns consists of adding, one at a time, meaningful control variables to equation (1), to get:

$$\ln Y = \beta_0 + \beta_1 L + \beta_2 E + \beta_3 E^2 + \beta_4 F + \beta_5 G + \beta_6 V + \varepsilon \qquad (2)$$

Equation (1) differentiates from equation (2) in that the latter includes an additional control variable (V). The empirical strategy is straightforward. It consists of observing the movements of the returns to literacy (β_1) after the inclusion of each additional control. If the included control is indeed an

[9] Reverse causality is considered a smaller issue in the literature, as literacy skills as measured by standardized tests are presumably determined before working life, such that high-paying jobs cannot influence them largely (Alderman et al., 1996, p. 40; Alderman et al., 1997, p. 121; Glewwe, 2002, p. 469; Hanushek/Zhang, 2009, p. 117). To confirm this notion, I subsequently removed individuals with the oldest ages from the sample and run model (1) after each removal. If it is the case that high-paying jobs steadily increase the literacy skills of individuals, one should observe that estimated returns to literacy decrease as a result of the removal of older individuals. Notice that this reasoning assumes that high-paying jobs are not overproportionately held by the younger employees of the sample. Estimations from this exercise, shown in the appendix, are rather stable and hence suggest that reverse causality do not drive the results of Table 2.

[10] Variables positively correlated with literacy and negatively influencing wages would be causing a downward bias. However, such variables are economically less plausible.

omitted variable upwardly biasing returns to literacy in equation (1), we should observe a decrease in returns to literacy (Oster, 2014, p. 2).

The richness of the STEP database allows for identifying economically meaningful control variables, that is, variables that are plausibly correlated with literacy skills and, in addition, might have an impact on hourly wages other than through literacy skills. As a result, the set of controls employed is significantly broader than those used in Hanushek et al. (2015). In that spirit, the literature on educational production functions and the competing theories of the human capital framework guided the selection of control variables.

There are two types of control variables I employ. The first type of controls are predetermined variables in the sense that they are determined previous to observed hourly wages. These are variables mostly related to family and formal education background: years of schooling, self-reported relative academic performance at school, kindergarten attendance, mother's education levels, socioeconomic status at age 15, household size at age 12, height, and number of economic shocks at age 12. Ideally, one would prefer control variables that are determined before literacy skills or, more generally, that are exogenous to the individual's literacy skills. That way, it is ensured that the control variable is not itself an outcome of literacy skills, avoiding the possibility that they might embody a channel through which literacy affects wages (Angrist and Pischke, 2008, pp. 64-68). In fact, all but the first three of the above-mentioned controls meet this stronger requirement. The second type of control variables are variables that are not necessarily predetermined to wages. These variables correspond to alternative indicators of human capital dimensions: self-reported Big Five personality traits, number of languages mastered, and the burden of chronic diseases in terms of temporary illness-related work unavailability. While useful, estimations including this type of controls should be interpreted cautiously, as controls coefficients might be suffering from reverse causality.

Years of schooling is the main control variable in this empirical analysis. Firstly, it is plausible to argue that years of schooling tend to positively correlate with literacy skills (Hanushek and Woessmann, 2008, p. 610). Secondly, the signaling (Spence, 1973, 2002), credentialism (Bills, 2003) and Marxist (Bowles and Gintis, 1975) approaches raise the relevance of years of schooling as a determinant of wages. These alternative theories develop mechanisms independent of cognitive production through which schooling affects wages. A decrease of the coefficient on literacy to insignificant levels after the inclusion of years of schooling would support these theories. Notice however, that such a finding would not rule out a positive role of literacy skills in wage determination. As higher literacy skills have been shown to increase schooling attendance, it could be the case that literacy skills still have an *indirect* wage impact through schooling attendance. In addition, the measurement error of literacy skills might be driving the empirical impact of schooling (Hanushek and Woessmann, 2008).

Other crucial control variables are mother's education levels and socioeconomic status at age 15. These are factors that have been reported as determinants of test scores, and in addition are pointed out by credentialist, Marxist, and noncognitive skills supporters as factors affecting wages. The possibility that children with highly educated parents and high socioeconomic status are able to invest in their cognitive skills production, and later on during their adult life have access to highly paid jobs mainly due to their socioeconomic background (Bills, 2003, p. 452), would bias returns to literacy upwardly. Moreover, highly educated parents might be more efficient at transferring noncognitive skills to their children, later rewarded in the labor market. Mother's education levels have been reported as an important factor influencing the health and nutrition of their children, which in turn might affect both literacy skills and wages of their children through a variety of channels (Klasen, 2002, pp. 351-352). By controlling for mother's education and socioeconomic status at age 15, many concerns related to these types of processes can be assessed.

Height has been reported as a proxy for cognitive and noncognitive skills that affect wages (see Case and Paxson (2006), Schick and Steckel (2010) and Sanchez (2013) for examples). The first association relates to the biological fact that well-nourishment at early ages fosters both the adult height and the brain development of individuals. If the literacy skill measure I employ is a good proxy for cognitive skills, the inclusion of height in the regression is redundant and it should not affect the coefficient on literacy through this channel. However, it has been argued that height also proxies for noncognitive skills such as authority and communication skills. Presumably, taller individuals overproportionately engage in sports at school, and as a result develop team-related skills such as discipline, confidence and leadership (Schick and Steckel, 2010, pp.3-5). If this reasoning is relevant, one should expect changes in the returns to literacy after the inclusion of height. As final remarks, height controls in wage equations might be more interesting in studies about developing countries, where undernutrition is a larger phenomenon. However, this control is more interpretable in country regressions than in cross-country ones, as it is less clear whether international height differences are majorly driven by malnourishment or genetic-related factors.

I also control for self-reported relative academic performance at school. If literacy skills are a good proxy for cognitive skills, this inclusion is redundant and should not cause a change in returns to literacy. Once again however, one could associate relative academic performance to noncognitive skills, in the sense that above-average achievement at school grants children with self-confidence.

Table 3. Returns to Literacy with predetermined and other control variables

Predetermined control variables	(1) Pooled	(2) Armenia	(3) Bolivia	(4) Colombia	(5) Georgia	(6) Ghana	(7) Kenya	(8) Vietnam
	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$
Baseline	0.285*** (0.015)	0.083*** (0.029)	0.317*** (0.043)	0.262*** (0.035)	0.171*** (0.034)	0.386*** (0.053)	0.271*** (0.048)	0.190*** (0.024)
Yrs schooling	0.085*** (0.019)	0.056** (0.028)	0.123*** (0.044)	0.080** (0.036)	0.120*** (0.034)	0.128** (0.057)	0.020 (0.054)	-0.001 (0.028)
Mother's education	0.256*** (0.020)	0.061** (0.029)	0.272*** (0.048)	0.246*** (0.050)	0.168*** (0.037)	0.347*** (0.056)	0.269*** (0.050)	0.180*** (0.028)
Status at age 15	0.267*** (0.018)	0.076*** (0.029)	0.280*** (0.041)	0.252*** (0.035)	0.166*** (0.034)	0.379*** (0.052)	0.261*** (0.048)	0.172*** (0.026)
HH size at age 12	0.285*** (0.018)	0.080*** (0.029)	0.319*** (0.043)	0.265*** (0.034)	0.168*** (0.034)	0.395*** (0.052)	0.270*** (0.048)	0.189*** (0.024)
Height	0.278*** (0.018)	0.081*** (0.029)	0.302*** (0.042)	0.259*** (0.037)	0.170*** (0.034)	0.386*** (0.056)	0.269*** (0.048)	0.188*** (0.024)
Kindergarten attendance	0.289*** (0.019)	-	0.319*** (0.043)	0.260*** (0.035)	0.169*** (0.034)	0.382*** (0.052)	0.269*** (0.048)	0.185*** (0.025)
Economic shocks until age 12	0.273*** (0.018)	0.080*** (0.029)	0.286*** (0.041)	0.258*** (0.032)	0.172*** (0.034)	0.380*** (0.052)	0.265*** (0.047)	0.181*** (0.024)
Academic performance	0.25*** (0.018)	0.066** (0.03)	0.296*** (0.043)	0.247*** (0.033)	0.140*** (0.036)	0.338*** (0.056)	0.254*** (0.053)	0.126*** (0.026)
Other controls								
Number of languages	0.268*** (0.018)	0.079*** (0.029)	0.310*** (0.041)	0.228*** (0.049)	0.137*** (0.036)	0.375*** (0.054)	0.258*** (0.050)	0.162*** (0.024)
Noncognitive skills	0.244*** (0.018)	0.083*** (0.03)	0.281*** (0.043)	0.231*** (0.037)	0.163*** (0.038)	0.254*** (0.065)	0.250*** (0.049)	0.151*** (0.025)
Chronic burden	0.285*** (0.018)	0.082*** (0.029)	0.322*** (0.042)	0.263*** (0.035)	0.172*** (0.034)	0.384*** (0.053)	0.269*** (0.048)	0.190*** (0.024)

Notes: OLS Regression results. Each row corresponds to a different estimated model. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for a constant, gender, experience and experience squared. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Mother's education, socioeconomic status at age 15, academic performance and economic shocks are categorical variables. Years of schooling, household size at age 12, height, chronic burden, the number of mastered languages and noncognitive skills measures enter the equation linearly. Detailed descriptions of each control variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

Kindergarten attendance is seen as strategically important in the formation of cognitive and noncognitive skills of children, and hence in wage determination (see Heckman (2006c) and Chetty et al. (2011) for examples), reason for which a dichotomous indicator for kindergarten attendance is used as a control in this framework. Motivated by the theory (Becker and Lewis (1974)) and empirics

(Kessler, 1991, pp. 415-416) of the quantity-quality trade-off of children, I also control for household size at age 12. The literature states that a larger household size results in reduced resource allocation per child, which has negative influences on socioeconomic outcomes. To the extent that these intra-household resource allocation effects on wages operate through channels other than literacy skills, the inclusion of this control variable should generate a downward movement in the coefficient on literacy. Similar to this reasoning, the number of economic shocks at age 12 is used as an additional control variable.

Finally, the analysis attempts to control for alternative measures of human capital. First, a set of 5 self-reported noncognitive skills measures known as the Big Five are together included in the estimation. These measures have gained relevance in empirical analysis for developed countries (Heckman, 2006b), which motivates their inclusion in this analysis.

A potentially important drawback of literacy skills assessments such as STEP and PIAAC is that they measure literacy skills in one language only, such that the number of languages one individual can speak is perhaps not properly represented by the literacy score. This neglect might be relevant as languages are an important dimension of human capital and particularly relevant in globalizing economies and their labor markets, as implied by Angrist and Lavy (1997). To tackle this concern, net hourly wages are also regressed on the number of languages mastered by each individual. Health indicators have been previously used as proxies for human capital (see Behrman (1999, pp. 2895-2901 for empirical review). In that spirit, I employ the burden of chronic disease in terms of temporary illness-related work unavailability in days as an additional regressor.

All cells of Table 3 show returns to literacy after controlling for predetermined and other control variables. Each row corresponds to a different estimated model. The first row is the baseline estimation replicated from Table 2, included in the table to ease the detection of coefficient movements. The following rows belong to estimations with the inclusion of additional control variables, which are indicated in the first column. Hence, coefficient movements should be detected vertically by comparing the coefficients in the first row with values of the following rows. The dependent variable, the regressor of interest and clustered standard errors are specified as in the previous estimation. All regressions control for experience, experience squared, gender and geographical location. The complete tables of each estimation are reported in the appendix.

The most important result to observe in Table 3 is that the returns to literacy remain highly significant, positive and relatively large for almost all estimations. That is a remarkable finding. Coefficient movements with respect to the baseline regression estimates have similar directions across samples, although the magnitude of the changes differs from sample to sample. In all cases, returns to literacy are more precisely estimated in the pooled regression. The precision of coefficient estimations are mostly unaffected by the inclusion of additional variables. Guided by changes in the pooled sample, I proceed to describe each estimation model, starting with those where coefficient movements are the largest. In Table 3, one can detect absolute changes easily. I make use of relative changes to capture the essence of coefficient movements and for comparison purposes.

The control variable generating the largest coefficient movement is years of schooling. In the pooled sample, returns to literacy drop 20 percentage points to reach a value of 8.5%, that is, a reduction of 70%. Nevertheless, the coefficient remains highly significant and of considerable magnitude. Conditioned on years of schooling, an increase of one standard deviation of literacy scores is associated with an 8.5% increase of hourly wages on average.

For the Armenian and Georgian sample, the reduction is relatively smaller, accounting for a drop of 33% and 30% to reach returns of 5.6% and 12%, respectively. Negative relative changes in Bolivia, Colombia and Ghana are double as much, moving within a range of 60% to 70%. Yet, coefficients remain large and significant, showing returns to literacy of 8%-13%. The point estimates for Ghana are similar to those reported in Glewwe (1996). For the exceptional cases of Kenya and Vietnam, the coefficient drop is virtually complete. The reductions are in order of 93% and 100% respectively, such that returns to literacy descend to insignificant levels. On the contrary, Boissiere et al. (1985) estimate returns to literacy of 19-22% for Kenya. However, their analysis encompasses individuals living in Nairobi exclusively, does not control for gender and makes use of a dichotomous rather than a continuous schooling variable.

An interesting turnover can be identified when comparing the estimated returns to literacy in Table 3 for the pooled sample with those reported in Hanushek et al. (2015) on developed countries.[11] Without controlling for years of schooling, returns to literacy seem to be higher for developing countries - 28.5% - than estimations for developed countries - 17.8% -, consistent with expected decreasing marginal returns. However, after conditioning on years of schooling, returns to literacy seem lower for developing countries - 8.5% - than those for developed countries - 10.1% -. Moreover, coefficients on years of schooling seem higher for developing countries - 8.5% - than for developed ones - 5.9% -.

Turning to the other empirical specifications of Table 3, it is evident that coefficient movements are significantly smaller than those observed with years of schooling. None of the remaining controls

[11] Both estimations control for the same variables and have similar sample restrictions. Estimations for developed countries correspond to returns to numeracy, although the authors mention that returns to literacy are very similar (Hanushek et al., 2015, pp. 15-17).

bring returns to literacy down to insignificant levels in any of the samples. After years of schooling, controls generating the largest relative changes are noncognitive skills, academic performance at school, and mother's education, which reduce estimated returns to literacy by 10%-15% for the pooled sample. The relative impact on the returns to literacy as a result of the inclusion of mother's education seems to be larger for the Armenian, Bolivian and Ghanaian samples, while the inclusion of a set of 5 noncognitive skills measures has a larger relative impact for the regressions of Ghana and Vietnam. In Ghana this reduction is of 34%. Self-reported socioeconomic status at age 15 and number of mastered languages reduce the coefficient on literacy by 5%-7% in the pooled sample. For the case of socioeconomic status at age 15, larger relative changes are observed for Armenia, Bolivia, and Vietnam, while number of mastered languages has a larger impact in the Georgian, Vietnamese and Colombian samples. The control for height, kindergarten attendance, household size at age 12, economic shocks until age 12 and chronic burden measure are causing minor to no changes in the returns to literacy in both the pooled and country samples.

Concerning the explanatory power of the control variables themselves, the coefficient on years of schooling stands out as the largest. The coefficient on schooling is highly significant, positive and large in all samples. This independent impact of schooling on hourly wages is slightly lower than the correspondent returns to literacy, with the exception of Colombia, where it is slightly higher, and Kenya and Vietnam, where years of schooling are several percentage points larger than the virtually inexistent returns to literacy. As expected, other control variables entering most of the estimations significantly positive are mother's education, socioeconomic status at age 15, number of mastered languages, academic performance at school, and stability and openness measures as a subset of non-cognitive skills.

Estimations with the control variables of Table 3 generally produce higher R-squared values. In all samples but the Armenian, upwardly R-squared movements[12] are larger when including years of schooling in the estimation. The pooled model can explain 33.8% of the variation in log net hourly wages. For country samples, R-squared values from estimations including years of schooling vary from 23.2% in Armenia to 40.1% in Colombia. Other control inclusions often increasing the explanatory power of the empirical model for the country samples are mother's education, socioeconomic status at age 15 and number of mastered languages, which raise R-squared values by 3-6 percentage points with respect to the baseline estimation. Some of these latter R-squared movements represent relative changes of more than 20%, which could be considered as very large (Oster, 2014, p. 29). The commented increases in R-squared values are relevant to the extent that they reduce unexplained

[12] See control coefficients and R-squared values in the appendix.

variation in the dependent variable. As a consequence of this reduction, there is less room for other potential omitted variables not controlled for, such that the likelihood of upward bias in returns to literacy gets further reduced, especially after considering that R-squared values derived from micro-econometric analysis are generally not too high (see Oster (2014) for a formalization of this rationale).

After describing the results reported in Table 3, three conclusions can be derived. First, the significantly positive effect of literacy skills on net hourly wages is robust to the inclusion of a meaningful and varied set of control variables for the pooled, Armenian, Bolivian, Colombian, Georgian, and Ghanaian sample. Kenya and Vietnam are exceptions to this general pattern indicating that returns to literacy can be very heterogeneous among developing countries. In general, and especially after taking potential attenuation bias into consideration, the findings reported in Table 3 support the cognitive skills approach of the human capital theory. Second, larger returns to cognitive skills measures for developing countries in comparison with estimations for developed countries are no longer observed after conditioning on years of schooling. Third, contrary to the prediction of the cognitive skills framework, years of schooling have a literacy-independent impact on hourly wages.

Acknowledging the important empirical role of years of schooling in the previous estimation, I exercise an even stronger robustness check, by again controlling for predetermined control variables but this time always including years of schooling in the estimation. Cells of Table 4 show returns to literacy and to schooling corresponding to this exercise. Only estimations causing relevant changes in the coefficients of interest are reported.[13] Each row reports estimations from a different model. The first row only controls for years of schooling, and is reproduced from Table 3 to ease coefficient comparisons. In other words, coefficient changes should be identified row-wise by comparing first row values with estimations of the following rows. Every other feature of the empirical estimation is identical to that of Table 3.

Returns shown in Table 4 indicate that after jointly controlling for years of schooling and mother's education, returns to literacy consistently remain significant and large for the pooled, Ghanaian, Colombian and Georgian sample. For the first two, estimated returns to literacy drop by less than 1 percentage point. For Colombia and Georgia, returns to literacy slightly increase. Conversely, estimated returns to literacy are no longer significant for the Armenian and Bolivian sample. Notice however, that this is mainly due to relatively large standard deviations. In the Armenian sample the drop is of one percentage point, from 5.6% to 4.6%. For Bolivia, estimated returns to literacy drop

[13] See appendix for extensions of Table 4 .

Table 4. Returns to Literacy and Years of schooling with predetermined control variables

Predetermined control variables	(1) Pooled		(2) Armenia		(3) Bolivia		(4) Colombia		(5) Georgia		(6) Ghana		(7) Kenya		(8) Vietnam	
	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$	$\beta_{Literacy}$	$\beta_{Schooling}$
Yrs schooling only	0.085*** (0.019)	0.084*** (0.004)	0.056** (0.028)	0.051*** (0.010)	0.123*** (0.044)	0.083*** (0.011)	0.080** (0.036)	0.092*** (0.010)	0.120*** (0.034)	0.082*** (0.010)	0.128** (0.057)	0.110*** (0.017)	0.020 (0.054)	0.105*** (0.013)	-0.001 (0.028	0.081*** (0.007)
Mother's educ	0.078*** (0.021)	0.083*** (0.004)	0.046 (0.029)	0.039*** (0.010)	0.086 (0.053)	0.087*** (0.014)	0.087* (0.049)	0.095*** (0.010)	0.125*** (0.037)	0.082*** (0.011)	0.119** (0.057)	0.105*** (0.018)	0.044 (0.055)	0.099*** (0.013)	0.000 (0.031)	0.084*** (0.007)

Notes: OLS Regression results. Each row corresponds to a different estimated model. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for a constant, years of schooling, gender, experience and experience squared. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Mother's education, socioeconomic status at age 15, academic performance and economic shocks are categorical variables. Years of schooling, household size at age 12, height, chronic burden, the number of mastered languages and noncognitive skills measures enter the equation linearly. Detailed descriptions of each control variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: STEP database.

less than 4 percentage points to a value of 8.6%. Although large, the coefficient is not precisely estimated. The standard deviation increases by 23%, raising multicollinearity issues. In addition, the Bolivian sample suffers a loss of statistical power. The number of observations decreased from 413 to 361, as individual information for mother's education is not always available. For this estimation, the 95% confidence intervals are very wide, containing values from -2% to 19%. Consistently, the estimated returns to literacy for Kenya and Vietnam remain insignificant.

Coefficients on years of schooling reported in Table 4 are positive, highly significant and large. The inclusion of mother's education as an additional regressor does not considerably affect their magnitude. For the pooled sample, an additional year of schooling is related with an 8.3% increase in net hourly wages, conditioned on literacy scores and mother's education. In the country samples, schooling coefficients vary from 3.9% in Armenia to 10.5% for the Ghanaian sample. Schooling coefficients are typically more precisely estimated than returns to literacy. This difference in estimation precision is responsible for schooling coefficients being significant in the Armenian and Bolivian sample, while coefficients on literacy scores are not, despite the fact that the latter are of larger size than the former. R-squared values further increase by 1 to 3 percentage points by the inclusion of mother's education as a control. As in estimations not controlling for years of schooling, other predetermined control variables entering most of the estimations significantly positive are socioeconomic status at age 15 and academic performance at school.

All together, the baseline estimation and the robustness checks suggest that estimated returns to literacy present *consistent heterogeneity in robustness.* Returns corresponding to the pooled, Colombian, Georgian and Ghanaian samples have consistently shown remarkable robustness. In the case of Armenia and Bolivia, estimations survive the first robustness check, but get challenged by the second strongest one, in part due to a loss in estimation precision related to multicollinearity and a loss of observations. The robustness of estimated returns to literacy for Kenya and Vietnam seem rather fragile, as returns promptly disappear in the first robustness check after including years of schooling in the empirical model. Granting more weight to the pooled sample estimations, and considering that the group of country samples with very robust estimations is the largest, these findings are considered to be sound support for the human capital theory. Nevertheless, it is important to acknowledge that the robustness checks employed in this analysis ease omitted variable concerns, yet they do not rule them out. To increase the reliability of causality inference, one would ideally employ more advanced identification strategies involving panel data techniques, twin data or randomized control trials (Hanushek et al., 2015), currently restrained due to data unavailability.

4.2.1. IV estimation as additional robustness check

Endogeneity issues threatening the validity of the baseline estimations have two counteracting sources. On the one hand, measurement error in literacy skills tends to downward bias the estimations. On the other hand, omitted variables, would tend to upwardly bias returns to literacy. As an additional robustness check, and in an attempt to reveal which of this endogeneity sources outweighs the other, I employ a 2SLS-IV estimation. The instrumental variables for literacy scores are school starting age and commuting time to primary school, classified in three categories: less than 30 minutes, 30 to 60 minutes, and more than 60 minutes. This information is collected retrospectively.

Later school entry ages and longer commuting time to primary school are expected to lower cognitive skills. Early formal education interventions have been reported as being successful at having long-lasting positive effects on cognitive skills. The idea behind is that the sooner the brain receives structured stimulus, the better it develops, easing cognitive production in later years (Heckman, 2008, pp. 308-313). For instance, a negative relationship between school starting age and IQ scores has been reported in an empirical study covering Norwegian workers (Black et al., 2008). Home-school distances have been previously used as factors raising the cost of schooling and hence negatively affecting its demanded quantity (Alderman et al., 1996, pp. 38-39). However, the mechanism at work here is a different one. Longer commuting times reduce the time availability for children to do homework and learn at home. In addition, longer commuting times to primary school might exhaust the brain, reducing its capability and appetite to absorb knowledge.

Table 5. 2SLS-IV estimation of Returns to Literacy

	(1) Second stage regression	(2) First Stage regression
Literacy	0.445^{***}	
	(0.168)	
Yrs schooling	0.047^{**}	0.109^{***}
	(0.019)	(0.004)
School starting age		-0.071^{***}
		(0.015)
Time to school		-0.067^{**}
		(0.027)
N	3661	3661
R^2	0.268	0.466

Notes: 2SLS-IV Regression results. Dependent variable for second stage regression: net log hourly wage in international dollars. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the pooled sample. The implemented literacy score values correspond to the arithmetic mean of the 10 plausible values. The second and first stage regressions control for a constant, experience, experience squared, years of schooling, socioeconomic status at age 15, a female dummy and country fixed effects. Years of schooling and socioeconomic status at age 15 enter the regression linearly. The instruments for literacy scores excluded from the second stage regression are school starting age and time needed for the home-school travel, both entering the first stage regression linearly. Detailed descriptions of the instrumental variables in appendix. Weak identification and overidentification tests support the validity of this instrument set. Each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: STEP database.

As shown in Table 5, IV estimations are restricted to the pooled sample to exploit larger variation in the instrumental variables. The estimation controls for socioeconomic status at age 15, gender, experience, experience squared and country fixed effects. Estimated returns to literacy are positive, highly significant and larger than OLS estimates. An increase in one standard deviation in literacy scores is associated with a 44.5% increase in net hourly wages. IV estimations for returns to cognitive skills in Behrman et al. (2009) for rural Guatemala show similar magnitudes. As it is expected, standard errors of the instrumented variable magnify. The coefficient on years of schooling halves but remains significant. This finding further supports the human capital theory. The first stage regression is depicted in the second column of Table 5. As expected, the coefficients on school starting age and time to school are negative and statistically significant, while years of schooling positively affect literacy scores.

The results suggest that attenuation bias dominates potential omitted variable bias of OLS estimations. In addition, the coefficient on years of schooling is reduced. This is in line with the reasoning that OLS coefficients on years of schooling are upwardly biased by the measurement error in literacy skills. With respect to the validity of the instruments, corresponding statistical tests perform well. The test of overidentifying restriction suggests that the exclusion restriction cannot be rejected by means of the J Hansen statistic. The corresponding p-Value is very large, taking the value of 73%. Hence, we cannot reject the null hypothesis that the instrument set is exogenous (Stock and Watson, 2007, pp. 443-449). With respect to the instruments relevance, comparing the cluster-robust Kleibergen-Paap rk Wald F statistic of 16.035 with the critical values of Stock and Yogo (2005), the estimation is able to reject already a 15% maximal IV size, that is, the rejection rate of the Wald test for the coefficient of interest is not far from the hypothesized 5% rejection rate, implying that the instruments are not weak (Baum et al., 2007, pp. 489-490).

Although the exclusion restriction cannot be rejected, economic considerations might challenge the presumed exogeneity of the chosen instrumental variables. For instance, school starting age and distance to school might well not be exogenous to certain family background characteristics of the child that could affect future hourly wages, such as parents' wealth. The inclusion of socioeconomic status at age 15 and years of schooling in the estimation eases such concerns. However, there are still plausible scenarios in which school starting age or distance to school can affect hourly wages, not explained by literacy skills, the socioeconomic background or the obtained school quantity of individuals. For instance, it could be the case that parents from children who lack noncognitive skills such as the capacity to stay still and pay attention, decide to send their children to school at a later age. Such deviations from average school starting ages are conceivable in developing countries, where compulsory starting ages are less binding. In any case, results of Table 5 are consistent with ex-ante economic and econometrical reasoning and standard instrumental validity tests confirm the appropriateness of the chosen instrumental variables, such that estimations can be interpreted as a useful additional robustness check.

4.3. Transmission channels

The findings reported in the previous sections support the role of cognitive skills as a fundamental determinant of hourly wages. An interesting inquiry that arises from this acknowledgment is: "what are the transmission channels behind the cognitive skills-wages relation?" To answer this question, I conceptualize literacy skills as the fundamental determinants of wages, and work activity categories by occupation, economic sector, formal/informal status and public/private job as the potential proximate causes to investigate. The idea behind this rationale is that individuals endowed with higher cognitive skills benefit from higher labor mobility across different types of work activities. As a result, they are able to access high-paying work activities.

In this vein, a reduction in the coefficient on literacy after the inclusion of one of the above-mentioned categorical variables would validate the categorical variable as a transmission channel (Hanushek et al., 2015, pp. 18-19). It is important to highlight that the inclusion of these variables is not intended to correct for potential omitted variable biases, but rather to uncover intermediate factors in the cognitive skills-wages relation. For that reason, these variables rather contain information on certain job characteristics. This clarification is important, as previous studies on returns to human capital measures combine omitted variable robustness checks and transmission channels investigations in the same analysis section, which can be misleading (Psacharopoulos and Patrinos, 2004, p. 116; Angrist and Pischke, 2008, pp. 64-68).

Table 5 shows the results of the transmission channels investigation. All cells show coefficients on literacy scores. All regressions control for years of schooling, experience, experience squared, gender and geographical location. Each row corresponds to a separate estimation. The included transmission variables are respectively indicated in the first column. Movements of the coefficient on literacy scores can be vertically identified by comparing the first row with subsequent rows. Further estimation specifications are consistent with the analysis shown in the previous section.

It is remarkable that coefficient movements across estimations are in general relatively low. Significance levels of coefficients vary depending on the sample and transmission factor investigated. Across samples, controlling for occupation fixed effects accounts for the largest coefficient movement. While for the pooled, Colombian and Georgian samples, the coefficient on literacy scores retain its statistical significance and is just slightly reduced by 1.6, 1 and 2.6 percentage points, in the Bolivian sample the coefficient roughly halves and drops to a value of 6.6%, only significant at the 10% level. For Ghana, the coefficient on literacy is reduced by almost 5 percentage points. It takes the value of 7.9%, yet it is not statistically different from 0, due to imprecise estimation. The coefficient on literacy turns insignificant in the Armenian sample by dropping 1.3 percentage points. Coefficients change less when controlling for economic sector fixed effects. Returns to literacy in the pooled sample drop by 1 percentage point, with similar drops among country sample estimations. Controlling for informality has a minor impact across samples. In the pooled sample, the drop is of 1 percentage point. Ghanaian estimations are an exception to this, as the coefficient drop is of almost 3 percentage points.

Table 6. Transmission channels controlling for years of schooling

Transmission channels	(1) Pooled	(2) Armenia	(3) Bolivia	(4) Colombia	(5) Georgia	(6) Ghana	(7) Kenya	(8) Vietnam
	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$
Yrs schooling only	0.085*** (0.019)	0.056** (0.028)	0.123*** (0.044)	0.080** (0.036)	0.120*** (0.034)	0.128** (0.057)	0.020 (0.054)	-0.001 (0.028)
Occupation	0.069*** (0.018)	0.043 (0.028)	0.066* (0.037)	0.070** (0.034)	0.094*** (0.034)	0.079 (0.056)	-0.001 (0.049)	0.006 (0.026)
Economic Sector	0.075*** (0.019)	0.056* (0.030)	0.111*** (0.041)	0.078** (0.037)	0.102*** (0.034)	0.123** (0.056)	-0.002 (0.052)	-0.003 (0.026)
Informal worker	0.076*** (0.019)	0.058** (0.028)	0.130*** (0.043)	0.082** (0.035)	0.104*** (0.034)	0.099* (0.057)	-0.004 (0.051)	-0.005 (0.028)
Public worker	0.084*** (0.019)	0.055* (0.028)	0.127*** (0.044)	0.086** (0.036)	0.121*** (0.034)	0.107* (0.058)	0.016 (0.050)	-0.001 (0.028)

Notes: OLS Regression results. Each row corresponds to a different estimated model. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for a constant, years of schooling, gender, experience and experience squared. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Occupation and economic sector are categorical variables, while informal and public worker are dichotomous variables. Years of schooling enters the estimation linearly Detailed descriptions for each transmission variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

R-squared values increase several percentage points with the inclusion of the transmission variables. For instance, the model with occupation fixed effects is able to explain 48% of the log net hourly wage variation for the Latin American countries. Turning to the significance of the transmission variables themselves, they all mostly enter the estimation significantly. As expected, informality negatively affects hourly wages, while the public worker dummy has a positive sign.

Interestingly, it seems that the access to certain types of occupations plays a meaningful role in the cognitive skills-wages relation. However, the finding that coefficients on literacy remain significant and relatively large after including occupation fixed effects suggests that the cognitive skills-earnings relation is able to explain hourly wage differences within occupations. The role of occupation access appears to be more important than the economic sector of the work activity. This finding seems plausible as occupation type is arguably more closely related to the cognitive demands of the job than the economic sector in which the worker operates. Both the type of occupation and the economic sector seem to be a more relevant transmission channel of cognitive skills than the formal/informal status of the job held. If informal and formal workers tend to concentrate in different occupations or economic sectors, then categories of the latter implicitly contain informality information. Notwithstanding, what strikes as unexpected is that the formal/informal status is not that closely involved in the cognitive skills-hourly wages relation. As a final remark, for the investigated relation, it seems less important whether the worker functions in the public or private sector.

5. Conclusion

The present work estimates the private returns to cognitive skills for a set of 7 developing countries. After conditioning for a set of economically meaningful control variables such as years of schooling, mother's education levels, socioeconomic status at age 15, height and noncognitive skills measures related to personality traits, among others, returns to literacy skills mostly remain positive and statistically significant. Conditioned on years of schooling, estimated returns for the pooled sample are lower than those reported in Hanushek et al. (2015) for developed countries, such that the assumption of marginal decreasing returns to cognitive skills is not empirically supported. Among the transmission channels investigated, increased labor mobility in terms of occupational choice stands out as an important proximate cause of hourly wage differences.

The analysis indicates considerable heterogeneity across countries. In the specification that controls for schooling attendance, point estimates for Bolivia, Georgia and Ghana are more than twice as large as those estimated for the Armenian sample, while estimated returns to literacy for Colombia stand at average levels. Returns to literacy in Kenya and Vietnam seem to be systematically lower than in the rest of the assessed countries. No clear decreasing pattern of returns to literacy skills by level of economic development can be detected. The heterogeneity of point estimates across countries suggests that while there is a tendency for reward of cognitive skills in most developing countries, labor markets from specific countries might systematically undervalue such skills. This statement exposes an exciting research agenda. Given that returns to cognitive skills measures seem to be a factor with considerable variation, it seems natural to investigate potential causes and consequences of those differences across economies.

Abstracting from uninvestigated general equilibrium effects, the estimated positive returns to cognitive skills support public interventions encouraging the accumulation of cognitive skills, as they are shown to yield economic benefits for individuals. Many policy tools involving the education and health sector fall under this umbrella. Unlike for the case of policy implications of estimations on returns to schooling, public investments aiming at improving schooling *quality* regain importance (Hanushek and Woessmann, 2008, pp. 658-659). Indeed, schooling quality has been suggested to become one of the building blocks of the post-2015 development agenda (UN, 2012, p. 34; UNESCO, 2014, p. 5). In addition to directly investing in the provision of human capital-enhancing education, public action might cost-effectively incentivize and efficiently redirect investments in education by enhancing access to labor market information on returns to cognitive skills (Rosenzweig, 2010, pp. 11-14; Jensen, 2010).

References

Alderman, Harold, 1996. The returns to endogenous human capital in Pakistan's rural wage labour market. *Oxford Bulletin of Economics and Statistics*, 58(1), pp. 29-55.

Alderman, Harold, Behrman, Jere R., Khan, Shahrukh, Ross, David R. and Sabot, R., 1997. The Income Gap in Cognitive Skills in Rural Pakistan. *Economic Development and Cultural Change*, 46(1), pp. 97-122.

Ammermueller, Andreas; Heijke, Hans; Woessmann, Ludger, 2005. Schooling quality in Eastern Europe: Educational production during transition. *Economics of Education Review*, 24(5), pp. 579-599.

Angrist, Joshua D. and Lavy, Victor, 1997. The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco. *Journal of Labor Economics*, 15(1), Part 2: Essays in Honor of Yoram Ben-Porath, pp. S48-S76.

Angrist, Joshua D. and Pischke, Jörn-Steffen, 2008. *Mostly Harmless Econometrics: An Empiristic's Companion*. Princeton: Princeton Univ. Press.

Baum, Christopher F.; Schaffer, Mark E. and Stillman, Steven, 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal*, 7(4), pp. 465-506.

Becker, Gary S., 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70(5), Part 2: Investment in Human Beings, pp. 9-49.

Becker, Gary S. and Lewis, H. Gregg, 1974. Interaction between Quantity and Quality of Children. In: Theodore W. Schultz, ed. 1974. *Economics of the Family: Marriage, Children, and Human Capital*, Chicago [u.a.]: University of Chicago Press, pp. 81-90.

Behrman, Jere R., 1999. Labor Markets in Developing Countries. In: Jelter, O. Ashen and Card, D., eds. 1999. *Handbook of Labor Economics*, Volume 3, Amsterdam: Elsevier Science B.V., pp. 2859-2920.

Behrman, Jere R.; Sengupta, Piyali and Todd, Petra, 2000. *Final Report: the impact of Progresa on achievement test scores in the first year*, International Food Policy Research Institute. Available at: <http://www.ifpri.org/publication/impact-progresa-achievement-test-scores-first-year> [accessed 18 February 2015].

Behrman, Jere R.; Hoddinott, John; Maluccio, John A. and Martorell, Reynaldo, 2009. Brains versus Brawn: *Labor Market Returns to Intellectual and Health Human Capital in a Poor Developing Country*. Middlebury College Economics Discussion Paper No. 0907. Available at: <http://www.middlebury.edu/~econ> [accessed 26 April 2015].

Ben-Porath, Yoram, 1967. The Production of Human Capital and the Life Cycle of Earnings. *The Journal of Political Economy*, 75(4), Part 1, pp. 352-365.

Berardi, Nicoletta, 2013. Social networks and wages in Senegal's labor market. *IZA Journal of Labor & Development 2013*, 2:3, pp. 1-26.

Berg, Ivar, 1971. Education and jobs: The great training robbery. Boston: Beacon Press.

Berndt, Ernst R., 1996. The Practice of Econometrics. Classic and Contemporary. Addison-Wesley.

Bills, David B., 2003. Credentials, Signals, and Screens: Explaining the Relationship between Schooling and Job. *Review of Educational Research*, 73(4), pp. 441-469.

Black, Sandra E.; Devereux, Paul J. and Salvanes, Kjell G, 2008. *Too Young to Leave the Nest: The Effects of School Starting Age.* IZA Discussion Paper Series 3452. Available at: <http://ftp.iza.org/dp3452.pdf> [accessed 20 March 2015].

Boissiere, M.; Knight, J. B. and Sabot, R. H., 1985. Earnings, Schooling, Ability, and Cognitive Skills. *American Economic Review*, 75(5), pp. 1016-1030.

Borjas, George J., 2013. *Labor economics.* Sixth Edition, International Edition. New York, NY: McGraw-Hill Irwin.

Bowles, Samuel and Gintis, Herbert, 1975. The Problem with Human Capital Theory--A Marxian Critique. *The American Economic Review*, 65(2), pp. 74-82.

Bowles, Samuel; Gintis, Herbert and Osborne, Melissa, 2001. The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature, 39(4), pp. 1137-1176.*

Brunello, Giorgio and Schlotter, Martin, 2011. *Non Cognitive Skills and Personality Traits : Labour Market Relevance and their Development in Education & Training Systems.* IZA Discussion Paper, No. 5743. Available at: <http://ftp.iza.org/dp5743.pdf> [accessed 16 April 2015].

Card, David, 1999.The Causal Effect of education on earnings. In: Ashenfelter, Orley C. et al., eds. 1999. *Handbook of Labor Economics,* Volume 3, part A., Amsterdam: Elsevier professional, pp. 1802-1859.

Case, Anne and Paxson, Christina, 2006. *Stature and Status: Height, Ability,and Labor Market Outcomes*. NBER Working Paper Series, Working Paper 12466. Available at: <http://www.nber.org/papers/w12466> [accessed 20 April 2015].

Chetty, Raj; Friedman, John N.; Hilger, Nathaniel; Saez, Emmanuel; Withmore Schanzenbach, Diane and Yagan, Danny, 2011. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4), pp. 1593-1660.

Cunha, Flavio; Heckman, James J.; Lochner, Lance and Masterov, Dimitry V, 2006. Interpreting the Evidence on Life Cycle Skill Formation. In: Hanushek, Eric A. and Welch, Finis, eds. 2006. *Handbook of the Economics of Education*, Volume 1, Amsterdam: Elsevier/North-Holland, pp. 697–812.

Educational Testing Services (ETS), 2014. *A guide to understanding the literacy assessment of the STEP Skills Measurement Survey.* Report, The World Bank.

Foster, Andrew D. and Rosenzweig, Mark R., 1996. Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution. *The American Economic Review*, 86(4),pp. 931-953.

Galor, Oded, 2005. From Stagnation to Growth: Unified Growth Theory. In: Aghion, P. and Durlauf, Steven N, eds. 2005. Handbook of Economic Growth. Volume 1A, Elsevier, pp. 172-293.

Gazzaniga, Michael S., 2009. *The Cognitive Neurosciences.* Fourth Edition, Cambridge: MIT PRESS.

Glewwe, Paul, 1996. The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development Economics*, 51, pp. 267-290.

Glewwe, Paul, 2002. Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. *Journal of Economic Literature,* 15, pp. 436–482.

Griliches, Zvi, 1969. Capital-Skill Complementarity. *The Review of Economics and Statistics,* 51(4), pp. 465-468.

Griliches, Zvi, 1977. Estimationg the Returns to Schooling: Some Econometric Problems. *Econometrica*, 45(1), pp. 1-22.

Hanushek, Eric A. and Luque, Javier A., 2003. Efficiency and equity in schools around the world. *Economics of Education Review*, 22, pp. 481-502.

Hanushek, E. A. and L. Woessmann, 2008. The Role of Cognitive Skills in Economic Development. Journal of Economic Literature 46, 607–668.

Hanushek, Eric A. and Zhang, Lei 2009. Quality-Consistent Estimates of International Schooling and Skill Gradients. *Journal of Human Capital*, 3(2), pp 107–143.

Hanushek, Eric A.; Schwerdt, Guido; Wiederhold, Simon and Woessmann, Ludger, 2015. Returns to Skills around the World: Evidence from PIAAC. *European Economic Review*, 73, pp. 103-130.

Harmon, Colm; Oosterbeek, Hessel and Walker, Ian, 2003. The Returns to Education: Microeconomics. *Journal of Economic Surveys*, 17(2), pp. 115-155.

Heckman, James J., 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), pp. 153-161.

Heckman, James J., 2008. Schools, Skills, and Synapses. *Economic Inquiry*, 46(3), pp. 289-324.

Heckman, James J.; Lochner, Lance J. and Todd, Petra E., 2006. Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In: Hanushek, Eric A and Welch, Finis, eds. 2006. *Handbook of the Economics of Education.* Volume 1, Amsterdam: Elsevier, pp. 310-451.

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

Heckman, James J., 2006c. Skill Formation and the Economics of Investing in Disadvantaged Children. *Science,* 312, pp. 1900-1902.

International Labour Organization (ILO), 2007. *Resolution Concerning Updating the International Standard Classification of Occupations.* Available at:<http://www.ilo.org/public/english/bureau/stat/ isco/ docs/resol08.pdf> [accessed 26 April 2015].

International Labour Organization (ILO), 2013. Cazes, Sandrine and Verick, Sher, eds. *Perspectives on Labour Economics for Development*. Geneva: International Labour Office.

International Labor Organization (ILO), 2013b. *19th International Conference of Labour Statisticians Geneva, 2–11 October 2013: Report of the Conference,* Report III. Available at: <http://www.ilo.org/stat/Publications/WCMS_234124/lang--en/index.htm> [accessed 26 April 2015].

Jensen, Robert, 2010. The (Perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics*, 125(2), pp. 515-548.

Jolliffe, Dean, 1998. Skills, Schooling, and Household Income in Ghana. *World Bank Economic Review,* 12(1), pp. 81-104.

Junker, Brian; Schofield, Lynne and Taylor, Lowell J., 2012. The Use of Cognitive Ability Measures as Explanatory Variables in Regression Analysis. *IZA Journal of Labor Economics*, 1:4, pp. 1-19.

Kessler, Daniel, 1991. Birth Order, Family Size, and Achievement: Family Structure and Wage Determination. *Journal of Labor Economics*, 9(4), pp. 413-426.

Klasen, Stephan, 2002. Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the Effect of Gender Inequality in Education on Economic Development. *World Bank Economic Review,* 16(3), pp. 345-373.

Lam, David and Levison, Deborah, 1992. Age, Experience, and Schooling: Decomposing Earnings Inequality in the United States and Brazil. *Sociological Inquiry*, 62(2), pp. 220–245.

Lavado, Pablo; Velarde, Luciana and Yamada, Gustavo, 2013. *Cognitive and Non-cognitive Skills: The Role of Latent Abilities and Wages on the Gender Wage Gap in Peru*. Working Paper 118, Young Lives. Available at: <http://r4d.dfid.gov.uk/pdf/outputs/Younglives/wp118-lavado-et-al.pdf> [accessed 26 April 2015].

Lee, Sangheon; McCann, Deindre and Messenger, Jon C., 2007. *Working Time around the world: Trends in working hours, laws and policies in a global comparative perspective*. New York et al.: Routledge.

Lemieux, Thomas, 2003. *The "Mincer Equation" Thirty Years after Schooling, Experience, and Earning.* Center for Labor Economics, University of California, Berkeley, Working Paper 62. Available at: <http://eml.berkeley.edu/~cle/wp/wp62.pdf> [accessed 26 April 2015].

Leuven, Edwin; Oosterbeek, Hessel and van Ophem, Hans, 2004. *Explaining international differences in male wage inequality by differences in demand and supply of skill.* University of Amsterdam, Working Paper. Available at: <http://www1.fee.uva.nl/scholar/wp/wp06-99.pdf> [accessed 26 April 2015].

Mankiw, Gregory N.; Romer, David and Weil, David N., 1992. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2), pp. 407-437.

Michaelowa, Katharina, 2001. Primary Education Quality in Francophone Sub-Saharan Africa: Determinants of Learning Achievement and Efficiency Considerations. *World Development*, 29(10), pp. 1699-1716.

Mincer, Jacob, 1970. The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach. *Journal of Economic Literature*, 8(1), pp. 1-26.

Mincer, Jacob A., 1974. Schooling and Earnings. The Human Capital Earnings Function. In: Mincer, Jacob A., ed. 1974. *Schooling, Experience, and Earnings.* New York [u.a.]: Columbia Univ. Press, pp. 41-63, 83-96.

Montenegro, Claudio E.; Patrinos, Harry Anthony. 2014. *Comparable estimates of returns to schooling around the world*. Policy Research working paper ; no. WPS 7020. Washington, DC

Neisser, Ulric; Boodoo, Gwzneth; Bouchard, Jr., Thomas J.; Boykin, A. Wade; Brody, Nathan; Ceci, Stephen J.; Halpern, Diane F.; Loehlin, John C.; Perloff, Robert; Sternberg, Robert J.; Urbina, Susana, 1996. Intelligence: Knowns and Unknowns. *American Psychologist*, 51(2), pp. 77-101.

Nelson, Richard R. and Phelps, Edmund S., 1966. Investment in Humans, Technological Diffusion, and Economic Growth. *The American Economic Review*, 56(1/2), pp. 69-75.

OECD, 2013. *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing.

OECD, 2014. OECD Stat database. Available at: <http://dx.doi.org/10.1787/data-00285-en> [accessed 13 April 2015].

Oster, Emily, 2014. *Unobservable Selection and Coefficient Stability: Theory and Evidence*. University of Chicago, Booth School of Business Working Paper. Available at: <http://brown.edu/research/ projects/oster/sites/brown.edu.research.projects.oster/files/uploads/selection.pdf> [accessed 26 April 2015].

Polachek, Solomon W. and Siebert, William S., 1993. *The Economics of Earnings.* Cambridge [u.a.]: Cambridge Univ. Press.

Ponce, Juan and Bedi, Arjun S., 2010. The impact of a cash transfer program on cognitive achievement: The Bono de Desarrollo Humano of Ecuador. *Economics of Education Review*, 29, pp. 116-125.

Pritchett, Lant, 2001. Where has all the education gone? *The World Economic Review*, 15(3), pp. 367-391.

Psacharopoulos, George and Patrinos, Harry Anthony, 2004. Returns to Investment in Education: A Further Update, *Education Economics*, 12(2), pp. 111-134.

Puhani, Patrick A., 2000. The Heckman Correction for Sample Selection and its Critique. *Journal of Economic Surveys*, 14(1), pp. 53-68.

Romer, Paul M., 1990. Human Capital and Growth: Theory and Evidence. *Carnegie- Rochester Conference Series on Public Policy,* 32, pp. 251-286.

Rosenzweig, Mark R., 1995. Why Are There Returns to Schooling? *The American Economic Review*, 85(2), pp. 153-158.

Rosenzweig, Mark R., 2010. *Microeconomic Approaches to Development: Schooling, Learning, and Growth*. Yale University, Center Discussion Paper No. 985. Available at: <http://www.econ.yale.edu/ growth_pdf/cdp985.pdf> [accessed 26 April 2015].

Sanchez, Alan, 2013. *The Structural Relationship between Nutrition, Cognitive and Non-cognitive Skills: Evidence from four developing countries*. Working Paper 111, Young Lives. Available at: <http://www.younglives.org.uk/publications/WP/the-structural-relationship between-nutrition-cognitive-and-non-cognitive-skills/wp111-sanchez-health-nutrition-skills> [accessed 26 April 2015].

Schick, Andreas and Steckel, Richard H., 2010. *Height as a proxy for cognitive and non-cognitive ability*. NBER Working Paper Series, Working Paper 16570. Available at: <http://www.nber.org/ papers/w16570> [accessed 23 April 2015].

Schreyer, Paul and Koechlin, Francette, 2003. Purchasing power parities – measurement and uses. *Statistics Briefs,* OECD, 3, Available at: <*www.oecd.org/std/prices-ppp/2078177.pdf*> [accessed 26 April 2015].

Schultz, Theodor W., 1961. Investment in Human Capital. *The American Economic Review,* 51(1), pp. 1-17.

Schuett, Florian, 2003. *The Importance of Human Capital for Economic Growth: Materialen des Wissenschaftsschwerpunktes „Globalisierung der Weltwirtschaft“*. Bremen: Universität Bremen, Institut für Weltwirtschaft und Internationales Management.

Soon, Lee-Ying, 1987. Self-Employment vs. Wage Employment: Estimation of Earnings Functions in LDCs. *Economics of Education Review,* 6(2), pp. 81-89.

Spence, Michael, 1973. Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), pp. 355-374.

Spence, Michael, 2002. Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review,* 92(3), pp. 434-459.

Stock, James H. and Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: D. W. K. Andrews, D.W.K. and Stock, James H. 2005. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, pp. 80–108.

Stock, James H. and Watson, Mark W., 2007. *Introduction to Econometrics.* Second Edition, International Edition. Boston, Mass. [u.a.]: Pearson/Addison-Wesley.

UNESCO, 2014. *Teaching and Learning: Achieving quality for all.*EFA Global Monitoring Report 2013/14. UNESCO Publishing. Available at: <https://en.unesco.org/gem-report/report/2014/teaching-and-learning-achieving-quality-all#sthash.QjOPoQ8Q.dpbs> [accessed 26 April 2015].

UN System Task Team on the Post-2015 UN Development Agenda (UN), 2012. *Realizing the Future We Want for All,* Report to the Secretary-General. Available at: <http://www.un.org/millenniumgoals/pdf/Post_2015_UNTTreport.pdf> [accessed 26 April 2015].

Verbeek, Marno, 2012. *A guide to modern Econometrics.* Fourth edition. Chichester: Wiley.

Valerio, Alexandria; Sanchez Puerta, Maria Laura; Tognatta, Namrata Raman; Monroy Taborda, Sebastian, 2016. *Are there skills payoffs in low- and middle-income countries ? Empirical evidence using STEP data*. Policy Research working paper; no. WPS 7879; WDR 2018 background paper. Washington, D.C.

Weiss, Andrew, 1995. Human Capital vs. Signaling: Explanations of Wages. *The Journal of Economic Perspectives*, 9(4), pp. 133-154.

Woessmann, Ludger, 2006. *International Evidence on Expenditure and Class Size: A Review*. University of Munich and Ifo Institute for Economic Research. Available at: <http://www.brookings.edu/gs/ brown/bpepconference/woessmann_paper.pdf> [accessed 26 April 2015].

World Bank, 2014. *STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills*. Social protection and labor Discussion Paper, No. 1421. Washington, DC: World Bank Group. Available at: <http://documents.worldbank.org/curated/en/2014/07/19894104/step-skills-measurement-surveys-innovative-tools-assessing-skills> [accessed 26 April 2015].

World Bank, 2015. *Purchasing Power Parities and the Real Size of World Economies: A comprehensive report of the 2011 International Comparison Program*. Available at: <http://www.gbv.de/dms/zbw/799817791.pdf> [accessed 26 March 2015].

World Bank, 2018. *World Development Report 2018*: *Learning torealize Education's Promise*. Washington, D.C.

Appendix

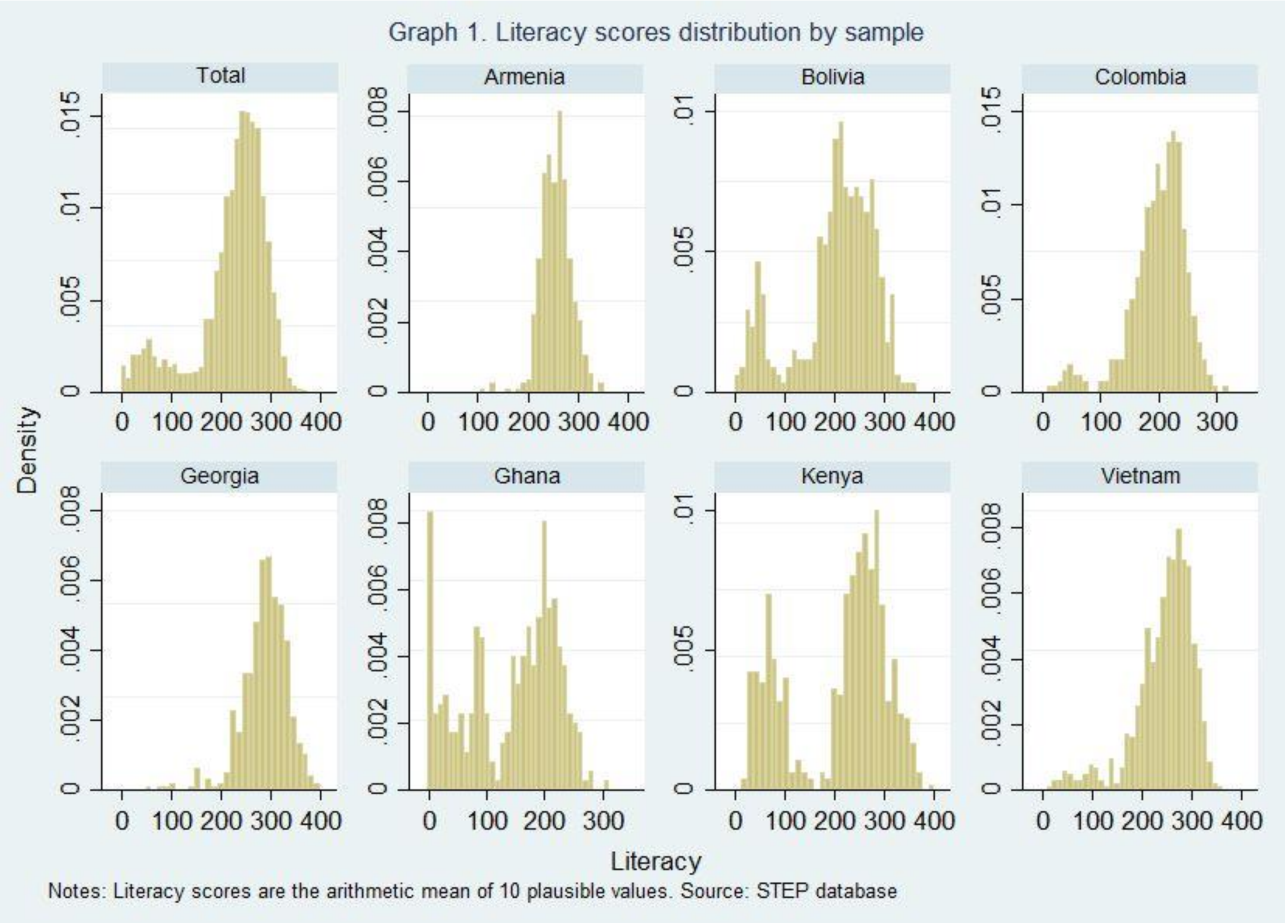

Table A1. Baseline regression: Returns to Literacy (extended)

	(1) Pooled	(2) Armenia	(3) Bolivia	(4) Colombia	(5) Georgia	(6) Ghana	(7) Kenya	(8) Vietnam
Literacy	0.285***	0.083***	0.317***	0.262***	0.171***	0.386***	0.271***	0.190***
	(0.018)	(0.029)	(0.043)	(0.035)	(0.034)	(0.053)	(0.048)	(0.024)
Experience	-0.032***	-0.001	-0.048**	-0.004	0.006	-0.072**	-0.101***	-0.014
	(0.007)	(0.012)	(0.020)	(0.013)	(0.019)	(0.031)	(0.024)	(0.012)
Experience^2	0.456***	-0.144	1.150***	0.001	-0.391	1.202*	1.719***	0.032
	(0.140)	(0.251)	(0.435)	(0.224)	(0.416)	(0.630)	(0.492)	(0.253)
Female	-0.233***	-0.404***	-0.284***	-0.296***	-0.290***	-0.150	0.062	-0.206***
	(0.025)	(0.048)	(0.074)	(0.050)	(0.068)	(0.107)	(0.079)	(0.042)
Constant	1.311***	6.531***	2.949***	8.587***	1.059***	1.400***	5.668***	10.412***
	(0.083)	(0.145)	(0.207)	(0.179)	(0.205)	(0.386)	(0.263)	(0.123)
Country fixed effects	Yes	No	No	No	No	No	No	No
Regional dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3727	568	413	496	492	351	541	866
R^2	0.240	0.194	0.217	0.253	0.183	0.241	0.173	0.165

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. Experience squared divided by 1000. The pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: STEP database.

Table A2. Returns to Literacy. Control variables: Years of schooling, Mother's education levels, Socioeconomic status at age 15

	(1) Pooled	(2) Pooled	(3) Pooled	(4) Armenia	(5) Armenia	(6) Armenia	(7) Bolivia	(8) Bolivia	(9) Bolivia	(10) Colombia	(11) Colombia	(12) Colombia
Literacy	0.085^{***}	0.256^{***}	0.267^{***}	0.056^{**}	0.061^{**}	0.076^{***}	0.123^{***}	0.272^{***}	0.280^{***}	0.080^{**}	0.246^{***}	0.252^{***}
	(0.019)	(0.020)	(0.018)	(0.028)	(0.029)	(0.029)	(0.044)	(0.048)	(0.041)	(0.036)	(0.050)	(0.035)
Yrs schooling	0.084^{***}			0.051^{***}			0.083^{***}			0.092^{***}		
	(0.004)			(0.010)			(0.011)			(0.010)		
Female	-0.219^{***}	-0.239^{***}	-0.249^{***}	-0.392^{***}	-0.410^{***}	-0.402^{***}	-0.294^{***}	-0.286^{***}	-0.328^{***}	-0.319^{***}	-0.267^{***}	-0.294^{***}
	(0.024)	(0.025)	(0.024)	(0.048)	(0.049)	(0.048)	(0.071)	(0.078)	(0.072)	(0.038)	(0.043)	(0.049)
1.mother's education		0.111^{**}			-0.528^{**}			0.271^{*}			0.025	
		(0.044)			(0.218)			(0.138)			(0.059)	
2.mother's education		0.258^{***}			-0.244			0.023			0.046	
		(0.051)			(0.195)			(0.143)			(0.133)	
3.mother's education		0.286^{***}			-0.191			0.165			0.357^{***}	
		(0.050)			(0.192)			(0.119)			(0.126)	
4.mother's education		0.340^{***}			-0.098			0.354^{***}			0.445^{**}	
		(0.059)			(0.194)			(0.134)			(0.200)	
5.mother's education		0.450^{***}			0.068			0.311^{**}			0.583	
		(0.061)			(0.195)			(0.152)			(0.447)	
6.mother's education		0.553^{***}			0.139						0.572^{***}	
		(0.076)			(0.199)						(0.161)	
Status at age 15 cat.	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Country fixed effects	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
Region fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3727	3449	3717	568	560	566	413	361	409	496	453	496
R^2	0.338	0.258	0.256	0.232	0.238	0.231	0.340	0.250	0.259	0.401	0.286	0.288

(continued)

	(13) Georgia	(14) Georgia	(15) Georgia	(16) Ghana	(17) Ghana	(18) Ghana	(19) Kenya	(20) Kenya	(21) Kenya	(22) Vietnam	(23) Vietnam	(24) Vietnam
Literacy	0.120***	0.168***	0.166***	0.128**	0.347***	0.379***	0.020	0.269***	0.261***	-0.001	0.180***	0.172***
	(0.034)	(0.037)	(0.034)	(0.057)	(0.056)	(0.052)	(0.054)	(0.050)	(0.048)	(0.028)	(0.028)	(0.026)
Yrs schooling	0.082***			0.110***			0.105***			0.081***		
	(0.010)			(0.017)			(0.013)			(0.007)		
Female	-0.310***	-0.305***	-0.309***	-0.178*	-0.151	-0.165	0.111	0.050	0.023	-0.157***	-0.189***	-0.216***
	(0.064)	(0.069)	(0.067)	(0.100)	(0.107)	(0.102)	(0.075)	(0.078)	(0.076)	(0.042)	(0.045)	(0.042)
1.mother's education		-0.370			0.181			0.354**			0.066	
		(0.344)			(0.279)			(0.152)			(0.067)	
2.mother's education		0.113			0.217			0.295**			0.208***	
		(0.274)			(0.153)			(0.138)			(0.074)	
3.mother's education		-0.034			0.403**			-0.192			0.278***	
		(0.259)			(0.196)			(0.187)			(0.083)	
4.mother's education								0.719**			0.295***	
								(0.303)			(0.105)	
5.mother's education		0.071			0.174			0.616***			0.348***	
		(0.258)			(0.346)			(0.209)			(0.116)	
6.mother's education		0.168									0.594	
		(0.260)									(0.479)	
Status at age 15 cat.	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Country fixed effects	No	No	No	No	No	No	No	No	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	492	482	491	351	317	350	541	509	539	866	767	866
R^2	0.264	0.205	0.202	0.354	0.246	0.268	0.295	0.222	0.220	0.297	0.185	0.192

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for experience, experience squared and a constant. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Mother's education and socioeconomic status at age 15 are categorical variables. Years of schooling enters the estimation linearly. Detailed descriptions of each control variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: STEP database.

Table A3. Returns to Literacy. Control variables: Height, academic performance and number of shocks

	(1) Pooled	(2) Pooled	(3) Pooled	(4) Armenia	(5) Armenia	(6) Armenia	(7) Bolivia	(8) Bolivia	(9) Bolivia	(10) Colombia	(11) Colombia	(12) Colombia
Literacy	0.278***	0.250***	0.273***	0.081***	0.066**	0.080***	0.302***	0.296***	0.286***	0.259***	0.247***	0.258***
	(0.018)	(0.018)	(0.017)	(0.029)	(0.030)	(0.029)	(0.042)	(0.043)	(0.041)	(0.037)	(0.033)	(0.032)
Height	0.006***			0.000			0.006**			0.005		
	(0.002)			(0.004)			(0.003)			(0.006)		
Female	-0.168***	-0.254***	-0.234***	-0.400***	-0.416***	-0.402***	-0.204***	-0.266***	-0.304***	-0.243***	-0.304***	-0.288***
	(0.029)	(0.025)	(0.024)	(0.068)	(0.048)	(0.049)	(0.078)	(0.077)	(0.072)	(0.093)	(0.051)	(0.051)
Academic performance	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Number of shocks	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Country fixed effects	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
Region fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3641	3668	3721	566	567	567	413	407	409	496	494	496
R^2	0.225	0.252	0.248	0.191	0.217	0.207	0.229	0.235	0.253	0.255	0.278	0.256

(continued)

	(13) Georgia	(14) Georgia	(15) Georgia	(16) Ghana	(17) Ghana	(18) Ghana	(19) Kenya	(20) Kenya	(21) Kenya	(22) Vietnam	(23) Vietnam	(24) Vietnam
Literacy	0.170^{***}	0.140^{***}	0.172^{***}	0.386^{***}	0.338^{***}	0.380^{***}	0.269^{***}	0.254^{***}	0.265^{***}	0.188^{***}	0.126^{***}	0.181^{***}
	(0.034)	(0.036)	(0.034)	(0.056)	(0.056)	(0.052)	(0.048)	(0.053)	(0.047)	(0.024)	(0.026)	(0.024)
Height	0.006			0.000			0.005			0.011^{***}		
	(0.005)			(0.005)			(0.004)			(0.004)		
Female	-0.228^{***}	-0.359^{***}	-0.294^{***}	-0.084	-0.128	-0.160	0.108	0.022	0.058	-0.095^{*}	-0.235^{***}	-0.209^{***}
	(0.088)	(0.072)	(0.068)	(0.118)	(0.108)	(0.106)	(0.084)	(0.082)	(0.078)	(0.056)	(0.041)	(0.042)
Academic performance	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Number of shocks	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Country fixed effects	No	No	No	No	No	No	No	No	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	492	492	492	282	339	350	526	516	541	866	853	866
R^2	0.185	0.212	0.186	0.254	0.246	0.244	0.176	0.194	0.181	0.172	0.216	0.178

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for experience, experience squared and a constant. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Academic performance and number of shocks are categorical variables. Height enters the estimation linearly. Detailed descriptions of each control variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

Table A4. Returns to Literacy. Control variable: Kindergarten attendance

	(1) Pooled	(2) Bolivia	(3) Colombia	(4) Georgia	(5) Ghana	(6) Kenya	(7) Vietnam
Literacy	0.289***	0.319***	0.260***	0.169***	0.382***	0.269***	0.185***
	(0.019)	(0.043)	(0.035)	(0.034)	(0.052)	(0.048)	(0.025)
Kindergarten	0.106***	0.011	0.077	0.100	0.207*	0.102	0.096*
	(0.030)	(0.079)	(0.052)	(0.076)	(0.114)	(0.084)	(0.055)
Female	-0.200***	-0.280***	-0.290***	-0.296***	-0.174*	0.061	-0.205***
	(0.028)	(0.076)	(0.051)	(0.068)	(0.106)	(0.078)	(0.042)
Country fixed effects	Yes	No	No	No	No	No	No
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
N	3134	407	493	491	347	532	864
R^2	0.256	0.222	0.252	0.186	0.255	0.174	0.169

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for experience, experience squared and a constant. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Kindergarten attendance is a dichotomous variable. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

Table A5. Returns to Literacy. Control variables: Chronic burden, Number of languages mastered, self-reported noncognitive skills

	(1) Pooled	(2) Pooled	(3) Pooled	(4) Armenia	(5) Armenia	(6) Armenia	(7) Bolivia	(8) Bolivia	(9) Bolivia	(10) Colombia	(11) Colombia	(12) Colombia
Literacy	0.285*** (0.018)	0.268*** (0.018)	0.244*** (0.018)	0.082*** (0.029)	0.079*** (0.029)	0.084*** (0.030)	0.322*** (0.042)	0.310*** (0.041)	0.293*** (0.043)	0.263*** (0.035)	0.228*** (0.049)	0.243*** (0.036)
Chronic burden	-0.021*** (0.007)			-0.022** (0.009)			-0.038*** (0.011)			0.002 (0.046)		
Female	-0.231*** (0.025)	-0.228*** (0.024)	-0.201*** (0.025)	-0.397*** (0.048)	-0.410*** (0.048)	-0.403*** (0.050)	-0.273*** (0.074)	-0.267*** (0.072)	-0.257*** (0.076)	-0.296*** (0.050)	-0.263*** (0.047)	-0.230*** (0.053)
Languages		0.150*** (0.021)			0.099*** (0.029)			0.165*** (0.058)			0.611*** (0.183)	
Extraversion			0.065*** (0.021)			0.060* (0.037)			0.041 (0.063)			0.022 (0.039)
Conscientiousness			-0.004 (0.026)			-0.033 (0.053)			0.008 (0.076)			-0.000 (0.059)
Openness			0.145*** (0.024)			0.111** (0.052)			0.019 (0.066)			0.147*** (0.040)
Stability			0.098*** (0.019)			0.075** (0.038)			0.136** (0.057)			0.105*** (0.033)
Agreeableness			0.009 (0.023)			-0.034 (0.048)			-0.032 (0.066)			0.030 (0.028)
Country fixed effects	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
Region fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3726	3725	3640	568	567	567	413	413	409	496	496	496
R^2	0.242	0.256	0.219	0.201	0.214	0.209	0.231	0.234	0.234	0.253	0.317	0.281

(continued)

	(13) Georgia	(14) Georgia	(15) Georgia	(16) Ghana	(17) Ghana	(18) Ghana	(19) Kenya	(20) Kenya	(21) Kenya	(22) Vietnam	(23) Vietnam	(24) Vietnam
Literacy	0.172***	0.137***	0.161***	0.384***	0.375***	0.254***	0.269***	0.258***	0.247***	0.190***	0.162***	0.156***
	(0.034)	(0.036)	(0.038)	(0.053)	(0.054)	(0.065)	(0.048)	(0.050)	(0.049)	(0.024)	(0.024)	(0.025)
Chronic burden	-0.019			0.021*			-0.055			-0.016		
	(0.015)			(0.011)			(0.044)			(0.015)		
Female	-0.297***	-0.326***	-0.314***	-0.157	-0.138	0.016	0.073	0.073	0.079	-0.206***	-0.180***	-0.138***
	(0.068)	(0.066)	(0.068)	(0.108)	(0.106)	(0.118)	(0.079)	(0.079)	(0.077)	(0.042)	(0.041)	(0.045)
Languages		0.207***			0.075			0.086			0.277***	
		(0.041)			(0.067)			(0.065)			(0.038)	
Extraversion			0.034			0.290***			0.061			0.068
			(0.059)			(0.098)			(0.063)			(0.044)
Conscientiousness			0.035			0.158			0.053			-0.039
			(0.072)			(0.127)			(0.085)			(0.048)
Openness			0.185***			0.060			0.299***			0.162***
			(0.059)			(0.107)			(0.086)			(0.042)
Stability			0.108**			-0.180*			0.066			0.147***
			(0.043)			(0.104)			(0.072)			(0.046)
Agreeableness			0.040			-0.053			0.061			0.000
			(0.060)			(0.085)			(0.082)			(0.046)
Country fixed effects	No	No	No	No	No	No	No	No	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	491	492	487	351	350	278	541	541	537	866	866	866
R2	0.186	0.222	0.220	0.242	0.243	0.199	0.175	0.176	0.214	0.166	0.214	0.201

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for experience, experience squared and a constant. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Chronic burden, number of languages mastered and the Big Five noncognitive skills measures enter the estimation linearly. Detailed descriptions of each control variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Source: STEP database.

Table A6. Returns to Literacy conditioned on Years of Schooling. Control variables: Mother's education, Status, Height

	(1) Pooled	(2) Pooled	(3) Pooled	(4) Armenia	(5) Armenia	(6) Armenia	(7) Bolivia	(8) Bolivia	(9) Bolivia	(10) Colombia	(11) Colombia	(12) Colombia
Literacy	0.078***	0.083***	0.081***	0.046	0.050*	0.056*	0.086	0.107**	0.123***	0.087*	0.086**	0.080**
	(0.021)	(0.019)	(0.019)	(0.029)	(0.029)	(0.029)	(0.053)	(0.042)	(0.044)	(0.049)	(0.035)	(0.036)
Yrs schooling	0.083***	0.082***	0.083***	0.039***	0.050***	0.051***	0.087***	0.083***	0.082***	0.095***	0.089***	0.092***
	(0.004)	(0.004)	(0.004)	(0.010)	(0.010)	(0.010)	(0.014)	(0.012)	(0.011)	(0.010)	(0.012)	(0.011)
1.mothers' education	0.016			-0.513**			0.050			-0.150***		
	(0.043)			(0.223)			(0.119)			(0.042)		
2.mother's education	0.138***			-0.247			-0.065			-0.143		
	(0.049)			(0.200)			(0.116)			(0.121)		
3.mother's education	0.122***			-0.215			0.013			0.066		
	(0.046)			(0.197)			(0.104)			(0.115)		
4.mother's education	0.173***			-0.150			0.214*			0.139		
	(0.054)			(0.200)			(0.114)			(0.138)		
5.mother's education	0.222***			-0.012			0.128			0.343		
	(0.057)			(0.201)			(0.133)			(0.426)		
6.mother's education	0.303***			-0.021						0.142		
	(0.073)			(0.208)						(0.145)		
Female	-0.225***	-0.228***	-0.184***	-0.398***	-0.387***	-0.398***	-0.306***	-0.333***	-0.272***	-0.323***	-0.315***	-0.310***
	(0.025)	(0.023)	(0.028)	(0.049)	(0.048)	(0.067)	(0.072)	(0.068)	(0.072)	(0.035)	(0.044)	(0.076)
Height			0.003*			-0.001			0.002			0.001
			(0.002)			(0.004)			(0.003)			(0.005)
Status at age 15 cat.	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Country fixed effects	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
Region fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3449	3717	3641	560	566	566	361	409	413	453	496	496
R^2	0.344	0.345	0.323	0.258	0.265	0.229	0.361	0.370	0.341	0.430	0.416	0.402

(continued)

	(13) Georgia	(14) Georgia	(15) Georgia	(16) Ghana	(17) Ghana	(18) Ghana	(19) Kenya	(20) Kenya	(21) Kenya	(22) Vietnam	(23) Vietnam	(24) Vietnam
Literacy	0.125***	0.117***	0.119***	0.119**	0.123**	0.110*	0.044	0.025	0.020	0.000	-0.006	-0.000
	(0.037)	(0.034)	(0.034)	(0.057)	(0.055)	(0.061)	(0.055)	(0.053)	(0.055)	(0.031)	(0.029)	(0.028)
Yrs schooling	0.082***	0.080***	0.082***	0.105***	0.115***	0.117***	0.097***	0.099***	0.105***	0.084***	0.079***	0.080***
	(0.011)	(0.010)	(0.010)	(0.018)	(0.016)	(0.020)	(0.013)	(0.013)	(0.013)	(0.007)	(0.007)	(0.007)
1.mother’s education	-0.273			0.042			0.294**			0.015		
	(0.273)			(0.256)			(0.145)			(0.063)		
2.mother’s education	0.153			0.040			0.246*			0.084		
	(0.237)			(0.145)			(0.129)			(0.069)		
3.mother’s education	-0.059			0.195			-0.251			0.103		
	(0.221)			(0.184)			(0.194)			(0.078)		
4.mother’s education							0.542			0.193*		
							(0.331)			(0.102)		
5.mother’s education	-0.038			-0.063			0.450**			0.119		
	(0.224)			(0.316)			(0.175)			(0.112)		
6.mother’s education	0.027									0.359		
	(0.224)									(0.471)		
Female	-0.310***	-0.320***	-0.247***	-0.191*	-0.174*	-0.118	0.109	0.079	0.158**	-0.139***	-0.165***	-0.095*
	(0.066)	(0.065)	(0.084)	(0.105)	(0.093)	(0.112)	(0.074)	(0.074)	(0.079)	(0.045)	(0.042)	(0.056)
Height			0.006			-0.001			0.005			0.006
			(0.005)			(0.005)			(0.004)			(0.004)
Status at age 15 cat.	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Country fixed effects	No	No	No	No	No	No	No	No	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	482	491	492	317	350	282	509	539	526	767	866	866
R^2	0.278	0.275	0.267	0.344	0.383	0.371	0.322	0.322	0.296	0.310	0.314	0.299

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to mean 0 and standard deviation 1 within corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for experience, experience squared and a constant. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Mother’s education and socioeconomic status at age 15 are categorical variables. Years of schooling and height enter the estimation linearly. Detailed descriptions of control variables in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

Table A7. Returns to Literacy conditioned on Years of Schooling. Control variables: Academic performance, Number of shocks, kindergarten attendance

	(1) Pooled	(2) Pooled	(3) Pooled	(4) Armenia	(5) Armenia	(6) Bolivia	(7) Bolivia	(18) Bolivia	(8) Colombia	(9) Colombia	(19) Colombia
Literacy	0.085***	0.081***	0.084***	0.055*	0.055*	0.119***	0.113**	0.132***	0.073**	0.078**	0.077**
	(0.020)	(0.019)	(0.021)	(0.029)	(0.028)	(0.045)	(0.044)	(0.044)	(0.036)	(0.035)	(0.037)
Yrs schooling	0.085***	0.083***	0.086***	0.044***	0.049***	0.082***	0.080***	0.083***	0.091***	0.092***	0.092***
	(0.004)	(0.004)	(0.004)	(0.012)	(0.010)	(0.012)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
Female	-0.233***	-0.221***	-0.185***	-0.395***	-0.391***	-0.280***	-0.310***	-0.291***	-0.323***	-0.317***	-0.313***
	(0.024)	(0.024)	(0.027)	(0.048)	(0.049)	(0.072)	(0.069)	(0.072)	(0.039)	(0.038)	(0.039)
Kindergarten			0.034					-0.067			0.045
			(0.027)					(0.066)			(0.039)
Academic performance	Yes	No	No	Yes	No	Yes	No	No	Yes	No	No
Number of shocks	No	Yes	No	No	Yes	No	Yes	No	No	Yes	No
Country fixed effects	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Region fixed effects	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3668	3721	3134	567	567	407	409	407	494	496	493
R^2	0.339	0.341	0.357	0.236	0.241	0.345	0.358	0.341	0.416	0.403	0.401

(continued)

	(10) Georgia	(11) Georgia	(20) Georgia	(12) Ghana	(13) Ghana	(21) Ghana	(14) Kenya	(15) Kenya	(22) Kenya	(16) Vietnam	(17) Vietnam	(23) Vietnam
Literacy	0.113***	0.121***	0.119***	0.101*	0.128**	0.130**	0.036	0.018	0.017	-0.005	-0.002	-0.003
	(0.035)	(0.034)	(0.034)	(0.061)	(0.057)	(0.057)	(0.055)	(0.054)	(0.055)	(0.028)	(0.028)	(0.029)
Yrs schooling	0.075***	0.081***	0.081***	0.122***	0.109***	0.107***	0.118***	0.104***	0.107***	0.076***	0.079***	0.080***
	(0.011)	(0.010)	(0.010)	(0.019)	(0.016)	(0.017)	(0.014)	(0.013)	(0.013)	(0.007)	(0.007)	(0.007)
Female	-0.331***	-0.312***	-0.314***	-0.196**	-0.181*	-0.191*	0.068	0.107	0.115	-0.172***	-0.160***	-0.156***
	(0.070)	(0.065)	(0.065)	(0.099)	(0.099)	(0.101)	(0.076)	(0.074)	(0.074)	(0.041)	(0.042)	(0.042)
Kindergarten			0.074			0.130			-0.015			0.060
			(0.075)			(0.107)			(0.076)			(0.051)
Academic performance	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No
Number of shocks	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Country fixed effects	No	No	No	No	No	No	No	No	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	492	492	491	339	350	347	516	541	532	853	866	864
R^2	0.268	0.265	0.266	0.350	0.353	0.361	0.330	0.300	0.300	0.310	0.304	0.298

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for experience, experience squared and a constant. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. Academic performance and number of shocks are categorical variables. Kindergarten attendance is a dichotomous variable. Years of schooling enters the estimation linearly. Detailed descriptions of each control variable in appendix. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

Table A8. Reverse causality Robustness check

age sample	(1) Pooled	(2) Pooled	(3) Armenia	(4) Armenia	(5) Bolivia	(6) Bolivia	(7) Colombia	(8) Colombia	(9) Georgia	(10) Georgia	(11) Ghana	(12) Ghana	(13) Kenya	(14) Kenya	(15) Vietnam	(16) Vietnam
	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$	$\beta_{Literacy}$
30-59	0.282***	0.083***	0.083***	0.056**	0.317***	0.123***	0.262***	0.080**	0.171***	0.120***	0.369***	0.118**	0.271***	0.020	0.190***	-0.001
	(0.018)	(0.019)	(0.029)	(0.028)	(0.043)	(0.044)	(0.035)	(0.036)	(0.034)	(0.034)	(0.055)	(0.058)	(0.048)	(0.054)	(0.024)	(0.028)
30-57	0.287***	0.087***	0.090***	0.063**	0.331***	0.138***	0.274***	0.084**	0.168***	0.114***	0.353***	0.120**	0.274***	0.021	0.203***	0.015
	(0.018)	(0.020)	(0.029)	(0.028)	(0.042)	(0.045)	(0.040)	(0.040)	(0.036)	(0.036)	(0.057)	(0.060)	(0.047)	(0.053)	(0.025)	(0.028)
30-54	0.280***	0.079***	0.086**	0.049	0.303***	0.105**	0.279***	0.087**	0.169***	0.114***	0.336***	0.101	0.279***	0.024	0.194***	0.009
	(0.019)	(0.020)	(0.034)	(0.032)	(0.043)	(0.048)	(0.044)	(0.042)	(0.036)	(0.035)	(0.059)	(0.064)	(0.048)	(0.053)	(0.025)	(0.028)
30-52	0.279***	0.083***	0.088**	0.054	0.311***	0.117**	0.277***	0.083*	0.173***	0.118***	0.336***	0.103	0.272***	0.024	0.182***	0.008
	(0.019)	(0.021)	(0.037)	(0.035)	(0.044)	(0.048)	(0.044)	(0.043)	(0.037)	(0.036)	(0.059)	(0.065)	(0.048)	(0.053)	(0.026)	(0.030)
30-49	0.271***	0.082***	0.066*	0.034	0.286***	0.082*	0.241***	0.080*	0.185***	0.128***	0.344***	0.124*	0.252***	0.007	0.173***	0.010
	(0.019)	(0.021)	(0.040)	(0.039)	(0.042)	(0.046)	(0.040)	(0.042)	(0.041)	(0.041)	(0.056)	(0.065)	(0.048)	(0.055)	(0.028)	(0.030)
30-47	0.266***	0.088***	0.075*	0.042	0.283***	0.076	0.239***	0.089**	0.186***	0.135***	0.339***	0.111	0.247***	0.021	0.163***	0.013
	(0.020)	(0.022)	(0.044)	(0.043)	(0.043)	(0.048)	(0.036)	(0.041)	(0.044)	(0.044)	(0.060)	(0.068)	(0.049)	(0.056)	(0.027)	(0.030)
30-44	0.269***	0.096***	0.078	0.040	0.280***	0.054	0.227***	0.068*	0.185***	0.137***	0.334***	0.133*	0.261***	0.042	0.151***	0.008
	(0.022)	(0.025)	(0.051)	(0.051)	(0.048)	(0.056)	(0.031)	(0.039)	(0.050)	(0.050)	(0.059)	(0.072)	(0.050)	(0.057)	(0.031)	(0.035)
Change 30-59:30-44	0.013	-0.013	0.005	0.016	0.037	0.069	0.035	0.012	-0.014	-0.017	0.035	-0.015	0.01	-0.022	0.039	-0.009

Notes: OLS Regression results. Dependent variable: net log hourly wage in local currency for country regressions and in international dollars for pooled regression. The 1st and 99th percentile of the net hourly wage distribution is trimmed. Sample: urban full-time wage workers aged 30-59. Literacy scores standardized to have mean 0 and standard deviation 1 within the corresponding samples. Estimations are the average of 10 estimations using a different plausible value each. All regressions control for a constant, gender, experience and experience squared. The second column of each sample controls additionally on years of schooling, which enters the estimation linearly. In addition, the pooled sample controls for country fixed effects, while the country regressions controls for regional fixed effects. In the pooled regression, each observation is weighted equally. Clustered standard errors in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01. Source: STEP database.

Table A9. List of variables STEP

Variable	Definition/Household survey questions/Scales
Regional controls	Armenia: 11 districts Bolivia: 4 districts Colombia: 21 municipalities Georgia: 11 regions Ghana: 10 regions Kenya: 4 regions Vietnam: 2 provinces
Clusters	For Armenia, Bolivia, Georgia, Kenya and Vietnam, each cluster consists of a census enumeration area with 15 observations per cluster. For Colombia, the clusters are 21 municipalities.
Years of schooling	The International Standard Classification of Education (ISCED) 1997 and 2011 scales are used to derive individual years of schooling. This variable represents the number of years of formal education corresponding to the highest level completed. In Bolivia, Colombia, Georgia and Vietnam, intervals of levels completed are narrow, such that one additional year often coincides with one level increase. In Armenia, Ghana and Kenya, however, the intervals of established levels in the education system are large, such that relevant years' variation gets lost. Hence, for these three countries, the actual years of formal education completed are taken. This variable captures primary, secondary, vocational and university education.
Mother's (Father's) education	Highest education level of the mother (ISCED 1997 scale): 0= Pre-school education, 1= Primary education, 2= Lower secondary education, 3= Upper secondary education, 4= Post-secondary non-teriary education , 5= Non-PhD tertiary education, 6=Phd.
Socioeconomic status at age 15	Survey question: Imagine a 10-step stairs where on the bottom, the FIRST step, stand the poorest people, and on the highest step, the TENTH, stand the richest. On which step do you think your family was when you were 15 years old?
Household size at age 12	Survey question: Which of your relatives usually lived in your household with you when you were 12 years old? (Father, Mother, Step-, Grand-, Brothers, Sisters, In-law-, Uncle and Aunt and their partners).
Kindergarten attendance	Survey question: Did you attend Pre-Kindergarten or Kindergarten?
Number of economic shocks	Survey question: Before you reached age 15, was the households financial situation significantly worsened because of death of household member, illness/serious accident of household member, family breakup/separation/divorce, alcohol problem/alcoholism or drug problem of household member, loss of employment for household member, bankruptcy or loss of family business, loss of crops, fire, drought, flood or other natural catastrophe, violence, theft, forced displacement or other? Scale: 0=no shocks, 1= 1 shock, 2= 2 or more shocks.
Academic performance	Survey question: From what you remember, how did you compare academically to your classmates in the highest grade you attended in primary or secondary school? Scale: 1= Below average, 2= Average, 3= Above average, 4= Excellent/among the best in class.
Number of mastered languages	Survey question: In which languages do you speak, and in which languages do you read and write, well enough to work in a job that requires that language?
Openness	Survey questions: Do you come up with ideas other people haven't thought of before? Are you very interested in learning new things? Do you enjoy beautiful things, like nature, art and music? Final measure corresponds to the average of the 3 responses on the following scale: 1=almost always, 2=most of the time, 3= some of the time, 4= almost never.
Conscientiousness	Survey questions: When doing a task, are you very careful? Do you work very well and quickly? Do you prefer relaxation more than hard work? Final measure corresponds to the average of the 3 responses on the following scale: 1=almost always, 2=most of the time, 3= some of the time, 4= almost never.

Extraversion	Survey questions: Are you talkative? Are you outgoing and sociable, do you make friends very easily? Do you keep your opinion to yourself? Final measure corresponds to the average of the 3 responses on the following scale: 1=almost always, 2=most of the time, 3= some of the time, 4= almost never.
Agreeableness	Survey questions: Do you forgive other people easily? Are you very polite to other people? Are you generous to other people with your time or money? Final measure corresponds to the average of the 3 responses on the following scale: 1=almost always, 2=most of the time, 3= some of the time, 4= almost never.
Stability	Survey questions: Are you relaxed during stressful situations? Do you tend to worry? Do you get nervous easily? Final measure corresponds to the average of the 3 responses on the following scale: 1=almost always, 2=most of the time, 3= some of the time, 4= almost never.
Chronic burden	Survey question: During the last 4 weeks, because of this chronic illness (diabetes, asthma, cancer, heart disease, high blood pressure, hepatitis, or other) how many days were you unable to carry out your usual activities?
Starting school age	Survey question: At what age did you start first course of primary? (age in completed years).
Commuting time to school	Survey question: When you were in Grade 1 or equivalent, how long did it take you to get to your primary school, by the usual method of transport you used? Scale: 1=Less than half an hour, 2= Between 0.5-1 hour, 3=More than an hour.
Occupation	The 1-digit ISCO Rev.8 occupation codes are used: 0=Armed forces occupations, 1= Managers, 2= Professionals, 3=Technicians and associate professionals, 4=Clerical support workers, 5=Service and sales workers, 6=Skilled agricultural, forestry and fishery workers, 7=Craft and related trades workers, 8=Plant and machine operators, and assemblers, 9=Elementary occupations.
Economic sector	The 2 digits ISIC revision 4 codes are used:1= Agriculture, forestry and fishing, 2= Mining and quarrying, 3= Manufacturing, 4= Electricity, gas, steam and air conditioning, 5= Water supply; sewerage, waste management and remediation activities, 6= Construction, 7= Wholesale and retail trade; repair of motor vehicles and motorcycles, 8= Transportation and storage, 9= Accommodation and food service activities, 10= Information and communication, 11= Financial and insurance activities, 12= Real estate activities, 13= Professional, scientific and technical activities, 14= Administrative and support service activities, 15= Public administration and defence; compulsory socialsecurity, 16= Education, 17= Human health and social work activities, 18= Arts, entertainment and recreation, 19= Other service activities, 20= Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use, 21= Activities of extraterritorial organizations and bodies.
Informal wage worker	Individuals who are currently working and reported not having social security or benefits are informal.
Social Network as job searching method	Survey question: What is the main method you used to find your work? Through public employment agency private employment agency university/school career office *social network (friends/relative/other)* media/internet Employer contacted you Contacted employer directly Started own business Job obtained after training/Apprenticeship with employer Other.

Source: STEP database, World Bank (2014), ILO (2007)