I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14289

Expecting Better? How Young People Form Their Earnings Expectations

Marta Favara
Paul Glewwe
Catherine Porter
Alan Sanchez

APRIL 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14289

Expecting Better? How Young People Form Their Earnings Expectations

Marta Favara
University of Oxford and IZA

Paul Glewwe
University of Minnesota

Catherine Porter
Lancaster University

Alan Sanchez
Group for the Analysis of Development

APRIL 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Expecting Better? How Young People Form Their Earnings Expectations[1]

Education choices are made based on the expected returns to schooling. If individuals are badly informed, they may make inefficient choices. We directly elicit young people's subjective expectations at the age of 14-15 about earnings under different educational scenarios and find these predict university enrolment by the age of 18-19. Females expect lower earnings than males, likely anticipating the reality of the labour market. Living in a poorer household, weaker numeric skills and lower self-efficacy are also associated with lower expected returns to education. Comparing expectations with the actual earnings from a nationally representative sample of individuals matched by sex, region and place of residence, we find that expectations for earnings upon completing secondary education closely match observed earnings, while there is a tendency to overestimate the returns to completing a university degree. These results hold for both males and females although with considerable variation across regions and population subgroups.

JEL Classification: I2, J22, J24
Keywords: subjective expectations, earning realizations, Young Lives, Peru

Corresponding author:
Marta Favara
Oxford Department of International Development
University of Oxford
3 Mansfield Road
OX1 3TB
Oxford, UK
E-mail: marta.favara@qeh.ox.ac.uk

[1] We would like to thank Maria Molina and Kristine Briones for the excellent research assistance.

1. Introduction

Higher education is seen as an important tool to promote social mobility and improve productivity of the future working population. Though there have been improvements in secondary enrolment (Glewwe and Muralidharan, 2015), in many low-and-middle income countries (LMICs) adolescents still drop out of secondary school at persistently high rates, particularly among the poor, which widens inequality (UIS and UNESCO GMR, 2016). Liquidity constraints, opportunity costs, and differences in schooling performance are largely understood as the main causes of early dropout (Glewwe and Muralidharan, 2015; Glewwe, 2004), which in turn has informed some policy interventions (e.g., Schultz, 2004). However, several studies also point to the role of information (e.g., Jensen, 2010; Nyugen, 2010). Although information is one factor in a multi-factorial problem, it has the advantage of being relatively inexpensive to fix.

Human capital theory assumes that people make decisions on their optimal level of schooling based on the monetary returns to education. This decision would require the individual to compare the (expected) costs and future outcomes from various choices, before selecting the option that maximizes their expected utility (e.g. Altonji, 1993). This poses the question as to whether people have enough information and an accurate perception of the role of education in determining future earnings (Betts, 1996). This type of question is difficult to answer in LMICs due to lack of data. Furthermore, while there is evidence about the role that subjective expectations play for education choices, we know surprisingly little about how people form these expectations. This is crucial to model people's choice of education particularly if individuals differ in the way in which they forecast earnings (Manski, 1993).

In this paper, we address three questions. First, what young people's subjective expectations about the return to education are, and to what extent they match earning realizations. Second, we investigate how individuals form their earnings expectations. Third, we investigate to what extent earnings expectations predict educational choices at the age of 18-19.

Using data from Young Lives Peru, a longitudinal study tracking a cohort of children from infancy to adolescence, we compare the earnings expectations of a 14-15 year-old cohort about what their earnings would be when they are aged 25, at different education levels, with the observed earnings in the population as captured by the Peru National Household Survey data from multiple years, matched at a local level and by gender (from here onwards, earnings realizations). In the Young Lives sample, earnings and employment expectations are elicited at age 14-15, when most of the children attend either 8th or 9th grade, two years away from the end of compulsory secondary education and the point at which students either terminate their education or decide to enrol in either university or vocational/technical education.

Our results suggest that earning expectations at the age of 14-15 predict university enrolment by the age of 18-19. Furthermore, young people's earnings expectations if they complete secondary education match closely the average earnings of 25-year olds from the same area

with the same level of education, while they tend to overestimate the earnings conditional on completed university. When investigating how young people form their earnings expectations, we found that living in a poorer household, having less educated parents, having lower education aspirations, poorer numeric skills and lower self-efficacy are all associated with lower earnings expectations for getting a university degree. Also, females have lower earnings expectations than males, possibly anticipating the reality in the labour market where observed earnings are considerably higher for males for both education scenarios.

This is one of only few papers investigating subjective expectations in low-and-middle-income countries (LMICs) in a non-experimental setting. Data on subjective expectations are quite scarce in LMICs (Delavande, 2014; Novella et al., 2018) and very few studies examine how earnings expectations predict educational choice and how expectations form. In fact, most of the empirical evidence on the returns to schooling come from studies using realized choices and earnings.

Our paper contributes to the recent and growing literature using subjective expectations data to understand decision making under uncertainty and limited information (Dominitz and Manski, 1996; Attanasio and Kaufman, 2014; Wiswall and Zafar, 2015; Rapoport and Thibout, 2018). This approach is motivated by evidence suggesting that individuals have biased expectations due to the limited information they have to support their thinking about returns to different education alternatives (Jensen, 2010). Expectation errors are important to the extent that they lead to suboptimal decisions regarding educational investments.

The paper is also related to the behavioural economics literature arguing that individuals can process only a limited amount of information at once and face impediments to their abilities of weighing the costs and benefits of every possible outcome of their decisions. This mental process behind any forward-looking decision is particularly costly for people living in poverty who face daily survival threats and have limited access to information (WDR, 2015).

In the last decade, Peru has made significant improvements in education enrolment, making it one of the best performers of its region. Between 2005 and 2015, the net enrolment rate (NER) and completion rates increased in all levels of education. The NER in tertiary education almost doubled during this period from 17% to 32% (ENAHO, 2015). Despite recent improvements, high school dropout rates are a significant problem and the recent pandemic might halt progress and potentially entrench existing inequalities for many young people. Persistently high dropout rates coexist with evidence suggesting that the return to secondary and tertiary education in Peru is generally high, although varying consistently across type of education institution and private and public institutions (see Lavado et al., 2014 for a review). Diaz et al. (2011) find that the effect size of an increase in 3 years of schooling in urban Peru is, on average, a 15% increase in earnings, conditional on skills (Diaz et al., 2011).

The remainder of this paper unfolds as follows: Section 2 reviews literature on the role of subjective expectations for educational choice. Section 3 presents the data used in this paper

and how expectations are measured in Young Lives. Section 4 describes the empirical approach used to investigate how young people form their expectations. Section 5 presents the main results and Section 6 concludes.

2. The role of expectations for educational choice in the literature

Robust evidence exists on the role that earnings expectations play for education choices. Attanasio and Kaufmann (2014), using survey data of Mexican junior and senior high school graduates containing information about labour market expectations of mothers and youth, find that expected returns and risk perceptions (of unemployment and earnings) are important determinants of the decision to enrol in senior high school and in college. Similarly, Arcidiacono et al., (2012) find that expected earnings play a significant role for the college major choice among students at Duke University, even after controlling for ability and career preferences. Hastings et al. (2016) highlights the importance to look not only at earnings expectations but also to costs expectations in investigating individual decisions on educational investments. They use administrative data and a large-scale survey of Chilean college applicants and students collecting expectations on earnings and costs at different institutions and majors. They find that while overestimating earnings does not lead to different matriculation rates, those who overestimate costs are less likely to matriculate in any degree program and are more likely to drop out. Kaufmann (2014) investigates the reason behind the difference in college enrolment among poor and rich young people in Mexico. She finds that poor individuals require higher expected returns to be induced to attend college than individuals do from rich families.

The evidence discussed suggests that providing information about the real return to schooling would potentially raise school enrolment, as found by Jensen (2010) in Dominican Republic and Nyugen (2010) in Madagascar. Nevertheless, in a number of studies in low-and middle-income countries, interventions combining information on the returns and costs of investing in post-secondary education together with information on requirements, eligibility and access to financial support, have been found to affect young people's expectations but without necessarily affecting their schooling choice (Bonilla et al., 2017 in Colombia; Busso et al., 2017 and Hastings et al., 2015 in Chile; Avitabile and de Hoyos (2018) in Mexico; see Novella et al., 2018 for an exhaustive review).

Similar evidence emerges in high-income country studies (Rapoport and Thibout 2018 in France; Wiswall and Zafar 2015 in the United States). Plausibly, the effectiveness of this type of policies depends on the context where they are implemented and the characteristics of the recipients. So for example, Bleemer and Zafar (2018), using a nationally representative sample in the United States, find that households from lower socioeconomic backgrounds are more likely to have biased beliefs about college costs and return and information interventions can reduce socioeconomic gaps in intended college attendance.

3. Data and descriptive statistics

Young Lives (YL) is a longitudinal study that follows two cohorts of children in Ethiopia, Peru, India (Andhra Pradesh and Telangana) and Vietnam: older cohorts born in 1994-1995 and tracked since ~ age 8 and younger cohorts born in 2001-2002 and tracked since ~ age 1. From here onwards, we focus on the Peru younger cohort sample (YC), the main cohort used for this analysis. The first study round was in 2002 and was followed by four subsequent in-person visits in 2006 (age 5), 2009 (age 8), 2013 (age 12) and 2016 (age 14-15). In 2020, when the younger cohort was aged between 18-19, a sixth survey round was conducted over the phone as an in-person visit was not possible due to the COVID-19 pandemic.

YL was developed as a longitudinal study of child poverty and the sampling design reflects that intent by oversampling poor households. Although YL is not intended to be representative of the country as a whole, because of the sampling procedure used, the YL sample for Peru has been found to optimally reflect the diversity of children and families in Peru, excluding the wealthiest 5% (Escobal and Flores, 2008).

Twenty clusters were randomly selected from the complete list of districts in Peru in 2002, excluding the wealthiest 5%. Each cluster was given a probability of being selected that was proportional to its population size. Then, within each selected cluster, families with children aged 6 to 18 months were selected to be part of the YC. The YC originally composed of 2,052 children and the attrition rate across all five in-person survey rounds is 9.8%. About 94% of the sample tracked in 2019/2020 participated in the 2020 phone survey, a very low attrition rates compared to similar follow-up phone surveys on cohort studies.[2]

Table 1 reports information on some individual and household level characteristics when the cohort was aged 14-15 and information about their education at age 14-15 and age 18-19. By age 14-15, 97% of the sample was school enrolled and the average participant was in the 8th grade In 2020, when we contacted the YL respondents for the last time, about 41% were in full-time education, 51% of them were in university, 35% were in technical institutes and 15% were still in secondary education.

[Insert Table 1]

A secondary source of data used for the analysis is the Peru National Households Survey (*Encuesta Nacional de Hogares*, ENAHO). ENAHO is a national multi-purpose survey carried out since 1995 with the aim to produce a wide range of indicators including data on the evolution of poverty, labour market indicators, incomes, well-being, and life-conditions of households in Peru. We use 2010-2016 ENAHO data for young adults aged 24-25 years old

[2] After 20 years, this represents 83% of the original sample.

and more specifically information about their gender, place of residence, education level, working status, and earnings realizations.

Using data from ENAHO, in Table 2 we report the employment rate and the (average) monthly earnings for individuals who completed secondary education and university, and the monetary return to university. The ENAHO sample is restricted to regions and provinces covered in the YL sample, for ease of comparison with earning expectations as discussed in the section 4. The average nominal annual earnings (from the main activity) are converted in monthly real earnings using 2016 as the year base.[3] The average monthly earnings for those who completed secondary education and those who complete university is respectively PEN 999 (above the minimum wage which was PEN 850 by the end of that year), which equates to 296 USD and PEN 1,332 which equates to USD 395.[4] The following patterns are observed. First, employment rates are significantly higher among young people with secondary education than among university graduates. Overall, males are more likely than females to be employed. Second, on average earning for those who completed university are significantly higher than the earning of those with secondary education. Third, males earn more than females regardless of the level of education completed. Fourth, there are substantial variations across regions in all dimensions, including returns to university (Table A1 in the Annex).

[Insert Table 2]

3.1 Eliciting subjective expectations in Young Lives

A survey module to capture self-earnings beliefs (or subjective earnings expectations) and employment probabilities was introduced in 2016 as part of Round 5 when the child was 14-15 years old, following the same methodology used by Dominitz and Manski (1996) and Attanasio and Kauffmann (2014).

The feasibility of eliciting subjective expectations is the object of some controversy in the research community. While this practice has been increasingly used in major surveys in developed countries (as for example in the UK Household Longitudinal study - Understanding Society), this type of data is very scarce in LMICs. The main concern is the respondent having difficulty in thinking in terms of subjective probability, particularly among illiterate people. A second concern is related to the long administration time. In both cases, the respondent would be less willing to respond to probabilistic expectations questions and the responses obtained might not be meaningful (Delavande, 2014; Delavande et al., 2017; Dominitz and Manski, 1997). More recently, innovative studies have experimented with modules providing

[3] The average earnings realizations by gender and area of residence (urban/rural) are reported at regional level in Table A1 in the Annex. Unfortunately, the sample size does not allow us to present the average earnings by gender and area of residence (rural/urban) at a lower geographical level than regional.

[4] The PEN/USD conversion is made using the average 2016 exchange rate: *PEN 1= USD 0.2965*

encouraging findings that suggest eliciting subjective expectations is feasible also in LMICs (Delavande and Kohler, 2009; McKenzie et al., 2007, Attanasio, 2009).

The 2016 Young Lives survey module included four questions on earnings expectations and employment probabilities. First, each participant was asked about the probability of working at age 25 conditional on having completed secondary education. More specifically, the interviewer asked the following questions: "*Assume that you finish secondary education, and that this is your highest schooling degree. From zero to ten, how certain are you that you will be working at the age of 25?"* Second, the interviewer asked the respondent what she believes the minimum and maximum she could earn per month would be conditional on having completed secondary education and working by age 25. More specifically, the interviewer asks: *"Now assume that you complete secondary education, and this is your highest schooling. Also, assume also that you find a job at age 25. What do you think is the minimum (maximum) amount you can earn per month at that age (age 25)?"*. Finally, since uncertainty about future earnings could play a role in schooling decisions, we incorporate this aspect in the survey by asking "*from zero to ten, what is the probability that your earnings at that age will be at least [x]*?" where [x] is the midpoint between the minimum and maximum amount elicited from the previous questions. The midpoint earning was automatically calculated by the software used to administer the survey and read to the respondent. The same questions were repeated for a second hypothetical scenario of completed university. The survey module was administered to all participants regardless of whether they were enrolled in school or not at the time of the interview.

Since the actual concept of probability may be difficult to grasp, every effort was made to facilitate the child understanding of the task. The module starts with a simple example of probabilities. Furthermore, a system of pebbles was used: the child was asked to select between 0 to 10 pebbles depending on how probable she thinks the situation is likely to happen and place them into a cup.

We use the information collected in the survey module described to define the probability density function of earnings and compute the main indicators used in the empirical analysis. To do so, we follow Attanasio and Kaufmann (2008) and we assume a triangular distribution. More specifically, we use the minimum reported earnings (y_m), the maximum reported earnings (y_M), and the probability mass p to the right of the mid-point of the support, with $p = Prob(y \geq (y_m + y_M)/2)$, to compute the first moment of the cumulative distribution function, i.e. the mean expected earnings (see Annex A3 for more details) (Attanasio and Kauffman, 2008).

Furthermore, we compute the return to university as the difference between the expected (log) earnings associated with university and secondary education. Finally, we used the information collected on the expected probability to be employed at the age of 25 for the two different schooling scenarios as indicators of perceived unemployment risk.

3.2 Validation of the subjective expectation data

There is no definitive way to assess to what extent respondents understood the question and provided accurate information. Nevertheless, looking at the prevalence of missing data, response patterns, including outliers, inconsistencies and bunching might be indicative of the quality of the data collected.

First, the Young Lives participants are willing to respond. We have very few cases where the respondent refused to answer the entire module (2%) or where one or more survey questions in the module is not answered (2.9%).

Second, respondents appear to understand the survey questions and the limited number of outliers and inconsistencies is encouraging. About 9% of the sample report the expected maximum or the minimum earnings to be more than 2.5 standard deviations above or below the mean, respectively. We consider them as outliers and drop them from the sample. Checking for inconsistencies, we did not find any case where the expected maximum earnings reported was smaller than the expected minimum amount for the same schooling level and only 2% of the cases where the opposite occurs (i.e. the minimum being greater than the maximum). Finally, we found only 1% of values outside the range 0-10 for those answers concerning probability. Overall, the small number of inconsistencies gives us confidence on the ability of the respondents to understand and answer the questions asked.

Third, a common fear is that respondents will round their responses values (0, 5, 10) instead of exploiting the full 0-10 scale (Dominitz and Manski, 1997). Figure 1 reports the complete frequency distribution of responses to the expected probability of working at age 25 conditioned on completed secondary education and university. Overall, respondents perceive high probabilities of events, being both distributions skewed toward the right and responses of zero chances being very uncommon. Most of them do not round their responses to the values (0, 5, 10). However, we find some bunching at values 5 and 10 for responses to the probability of working conditioned on completed secondary education (respectively 19% and 21%) and at values 10 for completed university (about 45%). We investigate whether the bunching is correlated on various attributes of respondents, with the aim to address the concern that bunching is prevalent among less wealthy and likely less educated respondents who might find the probability questions difficult to understand. Overall, we do not find any evidence of correlation between bunching at values (0, 5, 10) and the respondent's education level, the household welfare or parental education.

[Insert Figure 1]

3.3. What Young Lives young people expect: some descriptive statistics

Young people are asked about how much they believe they would earn at the age of 25, conditional on having completed secondary education (and university) as the maximum level

of education achieved and having a job. Overall, their earnings expectations are significantly higher for university than for secondary education. Conditional on completing secondary education, young people expect that at age 25 they will earn on average USD 236, which is just above the minimum wage prevalent up to mid-July (USD 222 per month). The average expected earnings for university at USD 631 per month are nearly 3 times the expectations for secondary education. This represents a return to university which is 167% higher than the expected earnings for secondary education.[5] Also, young people have better expectations about their chance to find a job as university graduates than if only completing secondary education (respectively with a probability of 85% and 64%).

Interestingly, earnings expectations for secondary education are about the same in urban and rural areas, but the earnings expectations for university are significantly higher in urban areas which leads to a higher expected return to university (173% versus 152% in rural areas). This might be explained by a higher demand for workers with university degrees in urban areas and/or by differences in the information available. In addition, males tend to have higher expectations than females. This gap in expectations is possibly anticipating the reality in the labour market, given that the earning realizations are higher for males, as further discussed in the next section.

4. Empirical strategy

Economic theory suggests that schooling choice largely depend on the relative market return of the different education alternatives available assuming that, at the time of schooling decision, young people know their ability, their taste for schooling and the present discounted life-cycle income that they would receive if working after each level of education. In context where the availability of labour market data is scarce, there might little information available on earnings. Furthermore, the decision to drop out of school is generally made at young ages when young people have not yet acquired any meaningful working experience and they might have a biased perception about their abilities, preferences and opportunities available to them.

Interactions with peers and family members play an important role in shaping young people's beliefs. They would draw inference about their own potential earnings by observing the incomes realized by members of the preceding generations or from others whom they can relate with (Manski, 1993). Growing up in poverty might hinder children's opportunities through a number of mechanisms: constraining opportunities available to them in the first place but also limiting their access to valuable information and role models, hinder their own aspirations and self-confidence. In addition, parental expectations for their children's future is likely to affect young people's self-beliefs too.

[5] The distribution of expected earnings and return to tertiary education is reported in Figure A1 in the Annex. Only 2% of the sample expect either a null return or a negative return to university.

Furthermore, existing evidence suggest that young people condition their expectations on their abilities (e.g. Willis and Rosen, 1979; Manski and Wise, 1983). More-able individuals might have higher expectations, being aware of their comparative advantage with respect to other students, and they might be more efficient at accessing and processing information about the population earnings distribution (Manski, 1993; Hastings et al., 2016). Furthermore, self-confidence is likely to influence their perception about their own abilities and ultimately their expectations about their academic and professional future (Bénabou and Tirole, 2002; Compte and Postlewaite, 2004).

The richness of the Young Lives datasets allows us to: first, investigate how subjective earning expectations varies across a broad number of observables characteristics at individual and household level; second to get a better understanding of how do they form their beliefs about the return to education; third, to investigate whether expectations predict educational choice.
To investigate the first question, we estimate $EE_{i,h,r}$, i.e. the expected (log) earnings of the individual *i* living in the household *h* in the region *r* for completed secondary education and completed university degree as the maximum education level. The first specification includes a core set of demographic and socio-economic characteristics:

$$EE_{i,h,r} = \boldsymbol{\gamma X_{i,h}} + +\boldsymbol{\mu_r} + \varepsilon_{ir} \quad (1)$$

More specifically, $X_{i,h}$ is a vector including the young people's age (in years), gender, household wealth (measured by the Young Lives wealth index, a composite measure of living standards, see Briones, 2017 for details), parental education (i.e. the highest educational attainment achieved by the two parents), having an older sibling and place of residency (living in rural or urban areas). Finally, $\boldsymbol{\mu_r}$ are region fixed effects control for (unobserved) labour market characteristics that could affect young people's expectations (including labour demand, labour supply, and earnings) common to all young respondents living in the same region. ε_{ir} is an error term, clustered at the region level to correct for within-region correlation. As for all the other specifications, all variables are either time invariant or contemporaneous with the dependent variable, measured at age 14-15.

A second specification includes additional controls measured in Round 5, at age 14-15. Exp_i captures parental expectations about the age when their child will become financially independent, Asp_i denotes young people's educational aspirations (it is a dummy equal to 1 if the young people aspires to (at least) complete university), W_i captures the child's working experience (it is a dummy equal to 1 if she/he is working in any paid activity for at least an hour per week), the vector Ed_i including the number of school grades completed by age 14-15 and whether the young person is studying in a public or private secondary school:

$$EE_{i,h,r} = \boldsymbol{\gamma X_{i,h}} + \partial_1 Exp_i + \partial_2 Asp_i + \partial_3 W_i + \emptyset \boldsymbol{Ed_i} + \boldsymbol{\mu_r} + \varepsilon_{ir} \quad (2)$$

In an extension of the model we control for the skills profile of the respondent.

$$EE_{i,h,r} = \boldsymbol{\gamma X_{i,h}} + \partial_1 Exp_i + \partial_2 Asp_i + \partial_3 W_i + \emptyset \boldsymbol{Ed_i} + \boldsymbol{\beta Cog_i} + \boldsymbol{\theta NCog_i} + \boldsymbol{\mu_r} + \varepsilon_{ir} \quad (3)$$

$\boldsymbol{Cog_i}$ is a vector capturing cognitive abilities at age 14-15 including numeracy and receptive vocabulary. Receptive vocabulary is measured in Young Lives using an adapted versions of the Peabody Picture Vocabulary Test (PPVT), a widely used test, administered between the ages of 5 and 14-15 years old (Dunn and Dunn, 1997). Numeracy skills are assessed using paper-based a mathematics test developed by Young Lives for the purposes of the survey. The math test incorporates questions at widely differing levels of difficulty, without restricting the level of difficulty assuming a specific grade.[6] The cognitive tests were administered to all children, regardless of whether or not they were attending school. This feature of the data avoids the selection problem that commonly arises when using school-based data.

$\boldsymbol{NCog_i}$ is a vector of non-cognitive skills including an indicator for general self-esteem and an indicator for generalized self-efficacy scales at age 14-15. The two indicators are constructed based respectively on an 8-item scale (Rosenberg, 1981) and a 10-item scale (Schwarzer and Jerusalem 1995). Self-esteem refers to an individuals' judgement of their own self-value or self-worth. Self-efficacy is defined as one's belief in one's capabilities to produce given attainments and to cope with adversity. It facilitates goal setting, effort investment, persistence in the face of barriers, and recovery from setbacks (Yorke and Ogando, 2018). Arguably, by including these two constructs the belief about one's earnings would better approximate the information the individual has on the population earning distribution, whilst controlling for the self-confidence element embedded in self-beliefs.

Finally, we add a control for the average earnings realization of a national representative 24-25 year old sample matched by sex, region and place of residence (urban/rural), RE_r :

$$EE_{i,h,r} = \boldsymbol{\gamma X_{i,h}} + \partial_1 Exp_i + \partial_2 Asp_i + \partial_3 W_i + \emptyset \boldsymbol{Ed_i}$$
$$+\boldsymbol{\beta Cog_i} + \boldsymbol{\theta NCog_i} + RE_r + \boldsymbol{\mu_r} + \varepsilon_{ir} \quad \textbf{(4)}$$

If respondents expect exactly the average earnings of her/his reference group, then we would expect the earnings realization to fully predict the earnings expectation, and all the other estimated parameters should statistically approximate to zero. Put differently, the estimated parameters would give as an indication on who is expecting better or worse than her/his own reference group once we have controlled for all the characteristics included in the specification above.

Finally, we investigate to what extent earnings expectations measured at age 14-15 predict school choice at the age of 18-19, the last time they were interviewed. More specifically, we estimate a linear probability model for the probability of being in university conditional on

[6] A validation of the psychometric properties of the PPVT and math scores can be found in Cueto and Leon (2012) and Cueto et al. (2009).

their earnings expectations and the individual and household level characteristics described above.

5. Results

5.1 Do young people expect better?

In this section we investigate the extent to which young people's earnings expectations for completed secondary and university education are aligned with the earnings of a national representative sample of 24-25-year-olds observed in the 2010-2016 ENAHO data. To improve precision, the match between earning expectations in the YL sample and earning realizations in the population is calculated separately by gender, region and area of residence (rural/urban). In Table 3, first, we compare the (average) earnings expectations and the (average) earnings realizations, separately for secondary education and university education.[7] Second, following Hastings et al. (2016), we define three groups: a first group of young people who "expect worse", i.e. whose (average) earnings expectations are at least 25% below the (average) earning realizations for the same level of education; similarly, a second group of young people who "expect better" i.e. whose (average) earnings expectations are at least 25% above the (average) earning realizations; and finally those who "expect correctly", i.e. whose earnings expectations are less than 25% above and below the (average) earning realizations. The last three columns of Table 3 report the percentage of those expecting worse, expecting correctly and expecting better, across different subgroups.

Males have significantly higher earnings expectations than females for both secondary and university. As for secondary education earnings, the proportion of females (and particularly urban females) expecting worse (41%) is significantly higher compared to the proportion of males (26%). No differences emerge between gender in the case of university.

[Insert Table 3]

The average expected earnings for university are substantially higher than the earnings observed in the population. Conversely, the average expected earnings for secondary education match the earnings realizations more closely. The distribution of the difference in earning expectations/realizations is represented in Figure 2.

[Insert Figure 2]

[7] The average earnings realizations at the regional level are reported in the Annex, Table A3, Table A4.

On average, the difference between earnings expectations and realizations for secondary education is close to zero.[8] The earnings expectations are only slightly below (USD 59) the earning realization (Table 3). However, the sample splits almost equally among those who expect worse (33%), those expecting correctly (37%) and those expecting better (30%; Table 3).

As for the difference between the earnings expectations conditional on completing university and the earning realizations of university graduates, on average young people expect in excess of USD 352 per month, a much bigger gap than for secondary education. [9] The majority of the respondents (57%) expect better while 32% of the sample expect correctly and only 11% expect worse (Table 3).

5.2 How young people form their earnings expectations?

In Table 4, we report the average expected earnings for the two education- attainment scenarios for different respondents identified by individual and household level observables characteristics measured at age 14-15.

[Insert Table 4]

Several interesting patterns emerge. First, as mentioned above, males have significantly higher earnings expectations than females for both secondary and university.

Second, young people living in rural areas, from the poorest households, with low educated mothers (none or only primary education completed) have similar expectations as their peers about secondary education earnings but tend to make significantly lower estimates of their earnings as university graduates, which leads to lower expected returns to university

Third, young people with better skills than the average (higher self-efficacy, self-esteem and higher numeracy and literacy skills at age 14-15), have higher earnings expectations compared to their peers.

Finally, young people who are still in education at the time of the interview (97% of the sample), as well as those who are in private secondary school expect a higher return to tertiary education than those who dropped out, while the two groups have similar earnings expectations for secondary education. Interestingly, young people who report to be working in paid activity

[8] Nevertheless, considerable variation across regions and subgroups emerges. In La Libertad, Amazons and Apurimac young people tend to overestimate the return to secondary education the greatest, while the opposite occurs in Puno, Arequipa and Lima.

[9] As for secondary education, the magnitude of the gap between earnings expectations and realizations varies substantially across regions, the return to university is overestimated in all regions, with very few exceptions, including Lambayeque and Loreto where urban males tend to underestimates their potential earnings realizations for university.

for at least an hour during the week, have higher expectations about what they could earn with completed secondary education.

Moving to a multivariate setting, Table 5 reports the results for the models specified in Section 5 for young people's earnings expectations if they complete secondary education (column 1-4) and if they complete university (columns 5-8).

[Insert Table 5]

Most of the variation in the earnings expectation for secondary education is due to differences across regions—captured by fixed effects (coefficients of regional fixed effects are omitted for simplicity)—and only gender and numeric skills are significantly correlated to them. In fact, numeric skills are positive associated with earnings expectations and females expect to earn about 24-26% less than males if they complete secondary education. The age at which the main caregiver expects him/her to become financially, independent is positively correlated with secondary education earnings (and the return to university). This result is not robust to the inclusion of the young people's (cognitive and non-cognitive) skills in the model.

As with secondary education, gender is also a robust predictor of university earnings expectations as females expect to earn about 22-25% less than males (Table 5, columns 5-7).

Furthermore, young people living in wealthier families and whose (at least one of the) parents completed higher education have higher earnings expectations. The role of parental education (only) is robust to the inclusion of the extensive set of controls.

Educational aspirations matter too. Those young people aspiring to complete university tend to have higher earnings expectations associated with the same education level, though of course these two variables are co-determined. The age at which the main caregiver expects him/her to become financially, independent is positively correlated with earnings expectations. However, this result is not robust to the inclusion of the young people's (cognitive and non-cognitive) skills in the model (result not reported).

Finally, education attainment (the number of grades completed by age 14-15) and having better numeric skills and higher self-efficacy are positively correlated with earnings expectations.

The findings discussed so far suggest that young people living in wealthier families, in urban areas, performing better at school, displaying better numeric skills, having higher aspiration and self-belief about their ability to succeed tend to have higher earnings expectations.

Interestingly, accounting for earnings realizations does not seems to alter the findings at all, and in fact earnings realizations do not seem to be correlated with earning expectations (Table 5, column 4 and column 8) despite the average deviation from reality in the sample being very small. This result will be further discussed below.

5.3 Do earnings expectations predict education choices?

As discussed in the previous section, expectations about the gain to completing secondary education are *on average* relatively accurate. This is not the case for university-level education as the average expectation is nearly twice the earnings realizations. Assuming that earnings expectations affect educational choices, our results suggest that a substantial fraction of the YL sample is likely to enrol in university or higher education.

In this section we investigate how accurate this hypothesis is by testing the extent to which earnings expectations predict enrolment in university at age 18-19 and more specifically, whether the young people who are most likely to enrol in a university are the ones who are more likely to overestimate the gains from doing so.

Columns 1-4 and columns 5-8 of Table 6 report the results for a linear probability model of enrolment in University when controlling respectively for the (log) earnings expectation and the (log) monetary return to university.

[Insert Table 6]

Earnings expectations and expected returns to university are positively associated with University enrolment. These results are robust to the inclusion of demographic and socio-economic characteristics (including sex, parental education, urban/rural, wealth index and having older siblings) and skills. However, once controlling for education trajectories (number of school grades completed by age 14-15 and having studied in a private secondary school), child's educational aspiration and parental expectations, the estimates for expectations are no longer significantly different from zero. These same factors explain differences in earnings expectations- which might explain why once included in the regression the association between earnings expectations and university enrolment is cancelled out. These results highlight the presence of persistent inequalities perpetuating along the life course and affecting school decisions which directly affect young people's opportunities and indirectly through their beliefs.

Among the other factors affecting university enrolment, sex and socio-demographic characteristics play an important role. Females, young people from wealthier families and whose parents have higher education are more likely to enrol in university. Education history matters for university enrolment, for example, young people who were attending private school at age 14-15 and/or by then have completed a higher number of school grades are more likely to reach university. Finally, skills matter too. Young people performing better in the cognitive tests and who have higher self-esteem at the age of 14-15 are more likely to be in university by the age of 18-19. Having an older sibling living in the household is the only variable negatively associated with university enrolment.

7. Conclusions

In this paper we investigate the expectations of a cohort of 15-years-olds in Peru about their earnings and labour market opportunities when they turn 25. We elicit their expectations using unique data from the Young Lives study. We investigate the extent to which these expectations reflect the earnings realizations of individuals aged 24-25 in the population (matched by region, sex, and area of residence). We find that their expectations for earnings conditional on completing secondary education match closely the earnings realizations of young people with that level of education. However, they tend to overestimate the earnings conditional on completing a university degree, although considerable variation in the difference between expectations and earnings across population subgroups emerges.

Using a multivariate approach to investigate the factors associated with earnings expectations, we found that females tend to have lower earnings expectations than males (likely anticipating the reality in the labour market). Furthermore, young people performing better at school, displaying better numeric skills, aspiring to complete university and self-efficacy tend to have higher earnings expectations. On top of these, living in wealthier families carries an additional premium in terms of expected earnings, which suggest structural wealth inequalities – possibly due to better connections (e.g. improving the likelihood of getting a good job).

The expected earnings do not seem to correlate with actual earnings realizations once controlling for a comprehensive set of individual and household level characteristics. The fact that the gender gap in expectations remains robust to the inclusion of earning realizations, even we account for their skills set, education history and parental background, might signal the presence of structural inequalities against women. Young women seem aware of the obstacles they will be facing to access rewarding jobs.

Finally, we investigate to what extent earnings expectations affect enrolment in university by the age of 18-19, the last time we interviewed the YL respondents. Our results suggest that earnings expectations predict enrolment in university conditional on socio-demographic characteristics and skills. Having unbiased beliefs about the return to education (conditional on the students' skills, school performance and available opportunities to them) might therefore be crucial for efficient educational choice.

Unfortunately, the data available do not permit one to distinguish between "population earnings beliefs" and "self-earnings beliefs". Clearly, there is a substantial difference between the two, and the process leading to the formation of such beliefs is likely to be distinct, as recognized by Beemer and Zafar (2018) and Wiswall and Zafar (2015). The main implication is that self-beliefs contrary to population beliefs, can only be partially attributed to variations in the information an individual has about the average salaries (Betts, 1996). On the other hand, population earnings beliefs can be directly validated by comparing those beliefs to the earnings realizations, but they do not necessarily form the basis of individuals' decision-making. Two individuals with the same population beliefs might make different choices if they have different beliefs about their own abilities and if they have different preferences and interests. It would be ideal to have information about both. Nevertheless, prior research has found a close connection between self-beliefs and population beliefs (Wiswall and Zafar, 2015).

The lack of information about population beliefs is not the only limitation of this paper. Two additional limitations can be identified. First, while this paper sheds some light on the factors associated with young people's expectations at a crucial age for their educational career, the results should not be interpreted in a casual way. Second, we do not have information about their expectations on the costs associated to each level of education, an important factor for educational choices. As Hastings et al. (2016) find, young people who overestimate costs are less likely to matriculate in any degree programme. Presumably, an overestimation of the associated cost might explain low investments in higher education, particularly for young people facing higher budget constraints, in presence of high earnings expectations.

References

Altonji, J G, 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain," *Journal of Labor Economics*, University of Chicago Press, vol. 11(1), pages 48-83, January.

Arcidiacono P.; Hotza V. J., Kanga S. 2012, "Modelling college major choices using elicited measures of expectations and counterfactuals"; *Journal of Econometrics*; Volume 166, Issue 1, January 2012, Pages 3-16

Attanasio, O. (2009), "Expectations and Perceptions in Developing Countries: Their Measurement and Their Use", *American Economic Review*, 99(2):87-92

Attanasio, O., Kaufmann, K., 2014. Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 2014, vol. 109, issue C, 203-216

Avitabile, C. and R. De Hoyos (2018)," The heterogeneous effect of information on student performance: Evidence from a randomized control trial in Mexico", *Journal of Development Economics*; Volume 135, November 2018, Pages 318-348

Beaman L., Duflo E., Pande R., Topalova P. (2012); Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India; *Science*, Vol. 335, Issue 6068, pp. 582-586

Benabou, R. and Tirole, J. [2002], 'Self-Confidence and Personal Motivation', *The Quarterly Journal of Economics*, 117(3), 871–915.

Betts, J., 1996. What do students know about wages? Evidence from a survey on undergraduates. *Journal of Human Resources/* 31, 27–56.

Bleemer, Z., B. Zafar (2018), "Intended college attendance: Evidence from an experiment on college returns and costs" */*, vol. 157, issue C, 184-211

Bonilla, L., N.L. Bottan and A. Ham (2017), "Information Policies and Higher Education Choices: Experimental Evidence from Colombia", Technical report, mimeo. Available at SSRN: *https://ssrn.com/abstract=2546835 or http://dx.doi.org/10.2139/ssrn.2546835*

Busso, M., T. Dinkelman, C. Martínez and D. Romero (2017), "The effects of financial aid and returns information in selective and less selective schools: Experimental evidence from Chile", *Labor Economics*, 45, 79-91.

Compte, O. and Postlewaite, A. [2004], 'Confidence-Enhanced Performance', *American Economic Review*, 94(5), 1536–1557.

Delavande, A., J. Lee and S. Menon (2017), "Eliciting Survival Expectations of the Elderly in Low-Income Countries: Evidence From India", *Demography*, volume 54, pages673–699(2017)

Delavande, A. (2014). "Probabilistic expectations in developing countries". *Annual Review of Economics, 6,* 1–20.

Delavande, A., & Kohler, H.-P. (2009). "Subjective expectations in the context of HIV/AIDS in Malawi". *Demographic Research, 20*(article 31), 817–875. doi:10.4054/DemRes.2009.20.31

Diaz, J.J., O. Arias, D. V. Tudela (2011), Does Perseverance Pay as Much as Being Smart?: The Returns to Cognitive and Non-cognitive Skills in urban Peru, *unpublished manuscript*.

Dominitz, Jeff and Charles Manski (f97) "Using Expectations Data to Study Subjective Income Expectations", *Journal of the American Statistical Association* 92(439): 855-67.

ENAHO, Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza, 2010-2016.

Escobal, J, E. Flores (2008), "An Assessment of the Young Lives Sampling Approach in Peru", *Young Lives Technical Note* N.3, Young Lives.

Galambos, N. L., Peterson, A. C., Richards, M., Gitelson. I. B. (1985). The Attitudes Towards Women Scale for Adolescents (AWSA): A study of reliability and validity. *Sex Roles 13* (5-6): 343- 356.

Glewwe, 2004. Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. *Journal of Economic Literature*, Vol. 40, No. 2 (Jun., 2002), pp. 436-482".

Glewwe, P. and Muralidharan, K. 2015. Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. *RISE Working Paper* 15/001.

Hastings, J., C.A. Neilson and S.D. Zimmerman (2015), "The Effects of Earnings Disclosure on College Enrollment Decisions", *NBER Working Paper* 21300.

Hastings, J.S., Neilson, C. A., Ramirez, A., Zimmerman, S. D. (2016) (Un)informed college and major choice: Evidence from linked survey and administrative data, *Economics of Education Review*, 51: 136-151.

Jacob, B.A., Wilder T. (2011). "Educational Expectations and Attainment." In Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low - Income Children, edited by Greg J. Duncan and Richard J. Murnane. New York, NY: Russell Sage Press.

Jensen R. (2010) "The (Perceived) Returns to Education and the Demand for Schooling", *The Quarterly Journal of Economics*, Volume 125, Issue 2, May 2010, Pages 515–548

Kaufmann M. K. (2014) Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns; *Quantitative Economics* 5 (2014), 583–630 1759-7331/20140583

Lavado, P., J. J. Martínez, G. Yamada (2014), Una promesa incumplida? La calidad de la educación superior universitaria y el subempleo profesional en el Perú, Tecnical note N° 2014-021, *Working Paper series*, Banco Central de Reserva del Peru.

Mckenzie, D, J. Gibson, S. Stillman (2007) "A Land of Milk and Honey with Streets Paved with Gold: Do Emigrants Have Over-Optimistic Expectations about Incomes Abroad?" *Journal of Development Economics* 102(2788) Follow journal

Manski, C. F. and D. Wise (1983) "College choice in America". Cambridge, MA: *Harvard University Press*, 1983. pp. 240

Manski, C. 1993 "Identification of Endogenous Social Effects: The Reflection Problem" *The Review of Economic Studies*, Volume 60, Issue 3, July 1993, Pages 531–542.

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from Madagascar. *Unpublished manuscript*, https://www.povertyactionlab.org/sites/default/files/documents/Nguyen%202008.pdf

Novella, R., A. Repetto, C. Robino and G. Rucci (2018), "Millennials en América Latina y el Caribe: ¿trabajar o estudiar? Una mirada regional" (with A. Repetto, C.Robino and G. Rucci) In: R. Novella, A. Repetto, C. Robino and G. Rucci (eds.), Millennials en América Latina y el Caribe: ¿trabajar o estudiar?, Washington DC: Inter-American Development Bank, 2018.

Rapoport B., Thibout C., (2018), Why do boys and girls make different educational choices? The influence of expected earnings and test scores, *Economics of Education Review*, Elsevier, 2018, 62, pp.205-229. ⟨10.1016/j.econedurev.2017.09.006

Schultz (2004). School subsidies for the poor: evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74(1), p. 199-250.

Stinebrickner, T., Stinebrickner, R., 2010. Math or science? The process of choosing a college major. *Working paper.*

Stinebrickner T., R. Stinebrickner, (2012). "Learning about Academic Ability and the College Dropout Decision," *Journal of Labor Economics*, University of Chicago Press, vol. 30(4), pages 707 - 748.

Streufert, P. (1991). "The Effect of Underclass Social Isolation on Schooling Choice." Institute for Research on *Poverty Discussion Paper* #954-91, University of Wisconsin-Madison.

UIS and UNESCO GMR (2016). "Leaving No One behind: How Far on the Way to Universal Primary and Secondary Education?"; UNESCO GMR.

Willis, R and S. Rosen (1979), "Education and Self-Selection", *Journal of Political Economy*, 1979, vol. 87, issue 5, S7-36

Wiswall, M., and Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital*, *9*(2), 117-169. https://doi.org/10.1086/681542

World Development Report (2015), World Development Report: Mind, Society and Behaviors, The World Bank, Washington DC, *https://www.worldbank.org/en/publication/wdr2015*

Zafar, B., 2008. College major choice and the gender gap. *Working paper*.

Zafar, B., 2011. How do college students form expectations? *Journal of Labor Economics*, 29 (2), 301–348.

Tables and Figures

Table 1. Characteristics of the Young Lives sample at age 14-15

	Mean	SD	N
Female	0.49	(0.500)	1860
Age in years	14.50	(0.500)	1858
Urban	0.74	(0.438)	1860
Having an older sibling	0.25	(0.433)	1860
Working in paid activity (at least 1hrs per week)	0.04	(0.194)	1857
Highest parental education			
Incomplete primary or less	0.23	(0.424)	1826
Complete primary to secondary	0.56	(0.495)	1826
Higher education	0.20	(0.397)	1826
Aspirations and expectations			
Caregiver's expected child's age when he/she becomes financially independent	24.35	(3.403)	1797
Child aspires to complete university	0.82	(0.382)	1818
Education			
Enrolled at age 14/15	0.97	(0.175)	1860
Highest grade completed by age 14/15	7.90	(1.241)	1860
In private school	0.18	(0.382)	1857
Enrolled at age 18/19	0.41	(0.013)	1552
% enrolled in university at age 18/19	0.26	(0.011)	1561
Region (% living in)			
Amazonas	0.048	(0.215)	1858
Ancash	0.095	(0.294)	1858
Apurimac	0.050	(0.218)	1858
Arequipa	0.048	(0.215)	1858
Ayacucho	0.088	(0.284)	1858
Cajamarca	0.052	(0.223)	1858
Callao	0.002	(0.046)	1858
Cusco	0.002	(0.046)	1858
Huanuco	0.031	(0.172)	1858
Ica	0.011	(0.106)	1858
Junin	0.047	(0.212)	1858
La Libertad	0.048	(0.214)	1858
Lambayeque	0.003	(0.057)	1858
Lima	0.181	(0.385)	1858
Loreto	0.001	(0.033)	1858
Madre De Dios	0.001	(0.023)	1858
Moquegua	0.001	(0.023)	1858
Pasco	0.003	(0.052)	1858
Piura	0.104	(0.306)	1858
Puno	0.041	(0.199)	1858

San Martin	0.091	(0.288)	1858
Tacna	0.002	(0.046)	1858
Tumbes	0.046	(0.210)	1858
Ucayali	0.002	(0.040)	1858

Note: Young Lives data, Younger cohort, Round 5 (2016).

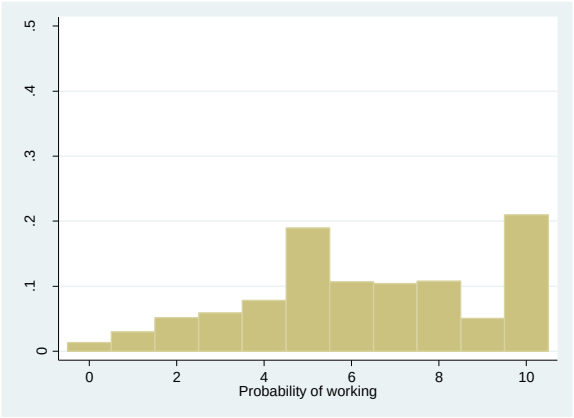
Table 2. Average monthly earnings and employment rate by level of education, gender and urban/rural (24-25-year olds)

						Males				Females			
		All	Rural	Urban	p-value	All	Rural	Urban	p-value	All	Rural	Urban	p-value
Earning realizations	Secondary education	296	214	306	0.000	345	242	359	0.000	220	150	227	0.000
	University	395	288	399	0.000	457	329	463	0.004	340	244	343	0.001
	Return to University	99	74	93		113	87	103		120	94	117	
		33%	35%	30%		33%	36%	29%		55%	62%	52%	
Employment rate	Secondary education	73%	83%	72%	0.000	87%	90%	86%	0.005	58%	71%	56%	0.000
	University	76%	75%	76%	0.548	79%	77%	79%	0.758	74%	74%	74%	0.488

Note: 2010-2016 ENAHO data; sample restricted to 24-25 years old and to Young Lives provinces only. Earnings expressed in USD using 2016 as year base (and an average PEN/USD exchange rate of PEN 1= USD 0.2965). The return to university (in USD) is computed as the difference between the (realized) earnings of those who completed a university degree and those who completed secondary education. The return to university (in %) is computed as a proportion of the earnings realizations for secondary education.

Figure 1. Probability of working at age 25 conditional on completed secondary education (Panel A) and University degree (Panel B)

Panel A



Panel B

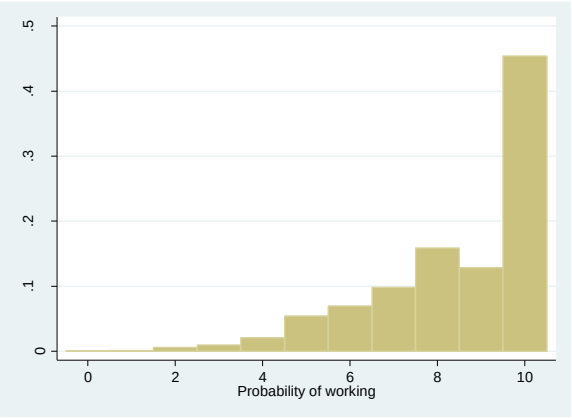


Table 3. Earnings expectations vs. earnings realizations at age 25 (in USD)

	(Average) Earnings Expectations	(Average) Earnings Realizations	% Expecting Worse	% Expecting Correctly	% Expecting Better
Secondary education					
All	237	296	33%	37%	30%
Female	206	220	41%	38%	21%
Male	268	345	26%	36%	38%
p-value	(0.000)	(0.000)	(0.000)	(0.512)	(0.000)
Rural	228	214	22%	28%	50%
Urban	240	306	42%	30%	28%
p-value	(0.315)	(0.000)	(0.000)	(0.291)	(0.000)
Female-Urban	208	227	52%	32%	16%
Male-Urban	271	359	32%	29%	39%
p-value	(0.000)	(0.000)	(0.000)	(0.218)	(0.000)
Female-Rural	197	150	24%	29%	47%
Male-Rural	257	242	20%	26%	54%
p-value	(0.000)	(0.000)	(0.287)	(0.554)	(0.157)
University					
All	747	395	11%	32%	57%
Female	652	340	11%	31%	58%
Male	839	457	10%	33%	56%
p-value	(0.000)	(0.000)	(0.660)	(0.296)	(0.476)
Rural	640	288	10%	18%	72%
Urban	781	399	13%	26%	61%
p-value	(0.011)	(0.000)	(0.217)	(0.001)	(0.000)
Female-Urban	676	343	13%	26%	61%
Male-Urban	883	463	12%	26%	62%
p-value	(0.001)	(0.000)	(0.820)	(0.953)	(0.835)
Female-Rural	576	244	10%	15%	74%
Male-Rural	701	329	10%	20%	70%
p-value	(0.041)	(0.003)	(0.843)	(0.215)	(0.356)

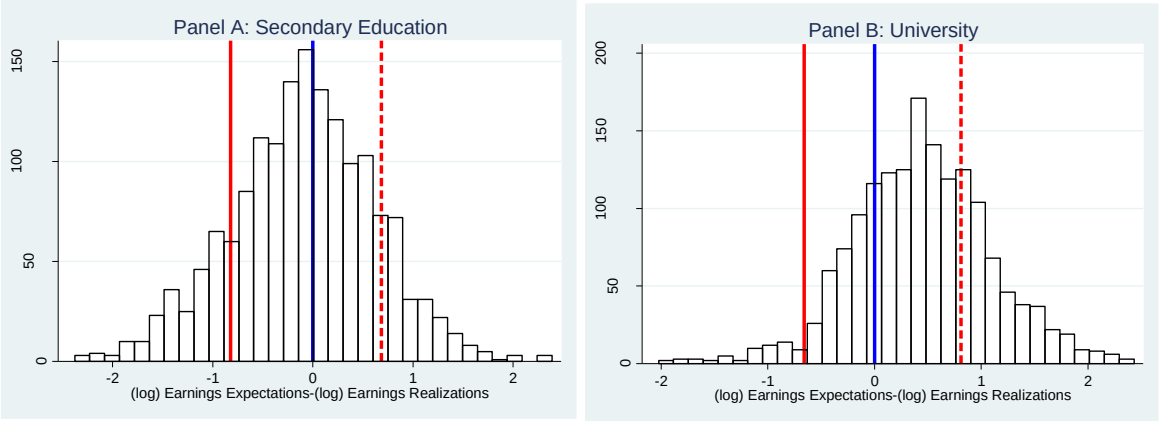
Note: Earnings realizations data: 2010-2016 ENAHO data; sample restricted to 24-25 years old and to Young Lives provinces only; population weight used. Earnings expectations: Young Lives data, Younger Cohort, Round 5 (2016). Earnings expressed in in USD using 2016 as year base (and an average PEN/USD exchange rate of PEN 1= USD 0.2965). "Expect worse (better)"=those whose (average) earnings expectations are at least 25% below (above) the (average) earning realizations for the same level of education; "expect correctly"=those whose earnings expectations are less than 25% above and below the (average) earning realizations. The match between earning expectations and realizations is done by gender, region and area of residence (i.e. rural/urban).

Table 4. Earnings expectations across different population subgroups (in USD)

	Expected earnings: complete Secondary educ.	Expected earnings: complete University	Expected return to University	Expected return to University (%)
All	236	631	395	167%
Sex				
Boys	263	696	432	164%
Girls	207***	563***	356***	172%
Parents' education				
None, primary	228	574	346	152%
Secondary +	238	650***	412***	173%
Location				
Rural	232	584	352	152%
Urban	237	646***	409***	173%
Household wealth				
Less poor	237	647	409	172%
Poorest	229	554***	325***	142%
Having older siblings				
No	237	631	394	166%
Yes	232	631	399	172%
Enrolment				
Out of education	198	402	204	103%
In education	237	636***	399***	168%
Working in paid activities				
No	234	632	398	170%
Yes	285***	594	308*	108%
Type of school				
Public	235	616	381	162%
Private	242	705***	464***	192%
Aspiring to university				
No	234	541	306	131%
Yes	237	649***	412***	174%
PPVT score				
High PPVT score	240	681	441	183%
Low PPVT score	230	569***	339***	147%
Math score				
High math score	251	706	456	182%
Low math score	223***	565***	342***	153%
Self-esteem				
High self-esteem	246	658	411	167%
Low self-esteem	229***	618***	385	156%
Self-efficacy				
High self-efficacy	248	702	454	183%
Low self-efficacy	226***	573***	347***	154%

*Note: Average values reported. Earnings expressed in in USD using 2016 as year base (and an average PEN/USD exchange rate of PEN 1= USD 0.2965). T-test for differences in mean across groups is performed and reported as *p-value < 0.1, **p-value < 0.05, ***p-value < 0.01. All time-variant characteristics are measured at age 15.*

Figure 2. Difference between earnings expectations and earnings realizations, by level of education



Note: Earnings realizations data: 2010-2016 ENAHO data; sample restricted to 24-25 years old and to Young Lives provinces only; population weight used. Earnings expectations: Young Lives data, Younger Cohort, Round 5 (2016). Earnings expressed in Peruvian Nuevo Sol (PEN). The figure shows the difference between the(log) earning expectations and the (log) earning realizations for complete secondary education (Panel A) and complete university (Panel B). The distribution on the left-hand side of the solid vertical red lines indicate those "expecting worse" (i.e. whose (average) earnings expectations are at least 25% below the (average) earning realizations for the same level of education). The distribution on the right-hand side of the dotted vertical red line indicated those who "expect worse" (i.e. whose (average) earnings expectations are at least 25% above the (average) earning realizations for the same level of education). The solid blue line in correspondence of zero indicates where the match between earning expectations and earnings realizations is perfect.

Table 5. OLS Regression results: log-earnings expectations

	(Log) Earning expectations: Secondary education				(Log) Earning expectations: University			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.257***	-0.254***	-0.242***	-0.254***	-0.230***	-0.247***	-0.222***	-0.220***
	(0.052)	(0.051)	(0.048)	(0.077)	(0.032)	(0.032)	(0.030)	(0.039)
Parent educ.: Complete primary/sec. education	0.023	0.015	0.006	0.006	0.067	0.033	0.016	0.017
	(0.038)	(0.040)	(0.041)	(0.041)	(0.049)	(0.048)	(0.048)	(0.047)
Parent educ.: Higher education	0.079	0.056	0.029	0.029	0.195***	0.124***	0.077*	0.079*
	(0.065)	(0.061)	(0.061)	(0.061)	(0.043)	(0.043)	(0.040)	(0.039)
Urban	-0.008	-0.001	-0.011	-0.008	0.023	0.010	-0.017	-0.015
	(0.050)	(0.050)	(0.046)	(0.046)	(0.062)	(0.059)	(0.054)	(0.057)
Wealth index: quintiles 2	0.070	0.058	0.064	0.064	0.052	0.021	0.021	0.019
	(0.044)	(0.044)	(0.043)	(0.043)	(0.058)	(0.056)	(0.054)	(0.055)
Wealth index: quintiles 3	-0.018	-0.036	-0.026	-0.026	0.146**	0.098	0.084	0.080
	(0.052)	(0.049)	(0.045)	(0.045)	(0.066)	(0.061)	(0.065)	(0.066)
Wealth index: quintiles 4	0.097	0.075	0.074	0.074	0.144**	0.083	0.057	0.052
	(0.060)	(0.054)	(0.055)	(0.055)	(0.064)	(0.059)	(0.061)	(0.062)
Wealth index: quintiles 5	-0.040	-0.064	-0.065	-0.065	0.116	0.045	0.019	0.015
	(0.088)	(0.082)	(0.081)	(0.081)	(0.088)	(0.089)	(0.089)	(0.091)
Having an older sibling	-0.007	0.001	0.001	0.001	-0.024	-0.010	-0.012	-0.012
	(0.040)	(0.042)	(0.043)	(0.043)	(0.052)	(0.053)	(0.052)	(0.052)
Work in paid activity		0.129	0.134	0.134		0.026	0.051	0.047
		(0.129)	(0.131)	(0.131)		(0.153)	(0.147)	(0.149)
Currently in education		0.125	0.113	0.114		0.107*	0.084	0.085
		(0.075)	(0.071)	(0.070)		(0.063)	(0.062)	(0.063)
Studying in private institution (age 14-15)		0.040	0.034	0.034		0.049*	0.027	0.028
		(0.033)	(0.032)	(0.032)		(0.025)	(0.029)	(0.030)
Parent's expectations: age financial independent		-0.002	-0.003	-0.003		0.006*	0.004	0.004
		(0.006)	(0.006)	(0.006)		(0.003)	(0.003)	(0.003)
Child: Aspire to university (age 14-15)		-0.006	-0.025	-0.025		0.109**	0.077*	0.074*
		(0.065)	(0.060)	(0.060)		(0.048)	(0.044)	(0.043)
Grade completed by age 14-15		0.037**	0.022	0.022		0.088***	0.057***	0.057***
		(0.017)	(0.019)	(0.019)		(0.016)	(0.017)	(0.017)
PPVT (std)			-0.024	-0.024			0.034	0.032
			(0.029)	(0.029)			(0.022)	(0.023)
Math (std)			0.072***	0.072***			0.071***	0.072***
			(0.016)	(0.016)			(0.014)	(0.014)
Self-esteem (std)			0.045	0.045			0.023	0.023
			(0.036)	(0.036)			(0.035)	(0.035)
Self-efficacy (std)			0.042	0.042			0.121***	0.121***
			(0.038)	(0.038)			(0.029)	(0.029)
(Log) Earning realization				-0.019				0.005
				(0.073)				(0.101)
Constant	6.578***	6.197***	6.384***	6.502***	7.542***	6.584***	6.983***	6.953***
	(0.083)	(0.217)	(0.229)	(0.602)	(0.076)	(0.148)	(0.155)	(0.684)
Observations	1,565	1,533	1,524	1,524	1,565	1,533	1,524	1,514
R-squared	0.060	0.064	0.076	0.076	0.081	0.106	0.134	0.133

Note: OLS estimates. All variables are either time invariant or measured at age 14-15. Parental education is the highest education achieved by the two parents (base category: primary incomplete or less). Wealth index base category: bottom quintiles. Regions fixed effects included in all specifications. Standard errors clustered at region level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. OLS regression results: being enrolled in University by age 18-19

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Log) Earning expectations	0.075***	0.063***	0.031**	0.014				
	(0.016)	(0.014)	(0.012)	(0.012)				
(Log) Return to University					0.053***	0.029*	0.010	0.009
					(0.015)	(0.014)	(0.009)	(0.015)
Female		0.094***	0.106***	0.088***		0.078***	0.099***	0.084***
		(0.020)	(0.021)	(0.022)		(0.021)	(0.022)	(0.023)
Parent educ.: Complete primary/sec. education		0.068**	0.032	0.006		0.070**	0.032	0.006
		(0.028)	(0.026)	(0.029)		(0.028)	(0.026)	(0.028)
Parent educ.: Higher education		0.293***	0.214***	0.167***		0.300***	0.216***	0.167***
		(0.040)	(0.043)	(0.044)		(0.040)	(0.042)	(0.044)
Urban		0.048	0.022	0.021		0.049	0.021	0.021
		(0.029)	(0.032)	(0.034)		(0.028)	(0.031)	(0.033)
Wealth index: quintiles 2		0.018	0.006	-0.015		0.022	0.008	-0.014
		(0.037)	(0.036)	(0.035)		(0.036)	(0.036)	(0.035)
Wealth index: quintiles 3		0.105***	0.071*	0.046		0.110***	0.073*	0.047
		(0.034)	(0.036)	(0.036)		(0.034)	(0.037)	(0.036)
Wealth index: quintiles 4		0.137***	0.096**	0.061*		0.146***	0.099**	0.062*
		(0.038)	(0.038)	(0.033)		(0.038)	(0.038)	(0.033)
Wealth index: quintiles 5		0.272***	0.228***	0.178***		0.276***	0.228***	0.178***
		(0.044)	(0.040)	(0.038)		(0.044)	(0.040)	(0.037)
Having an older sibling		-0.050**	-0.047**	-0.039*		-0.051**	-0.047**	-0.039*
		(0.021)	(0.020)	(0.022)		(0.021)	(0.020)	(0.022)
PPVT (std)			0.058***	0.036***			0.058***	0.036***
			(0.014)	(0.012)			(0.014)	(0.011)
Math (std)			0.065***	0.055***			0.068***	0.056***
			(0.014)	(0.017)			(0.014)	(0.017)
Self-esteem (std)			0.036*	0.043**			0.037*	0.044**
			(0.019)	(0.019)			(0.019)	(0.019)
Self-efficacy (std)			0.015	0.004			0.018	0.005
			(0.030)	(0.030)			(0.030)	(0.030)
Studying in private institution (age 14-5)				0.117***				0.117***
				(0.022)				(0.022)
Parent's expectations: age financial independent				0.004				0.004
				(0.003)				(0.003)
Child: Aspire to university (age 14-15)				0.033				0.034
				(0.039)				(0.039)
Grade completed by age 14-15				0.067***				0.068***
				(0.019)				(0.019)
Constant	-0.413***	-0.521***	-0.192**	-0.682***	0.105***	-0.069*	0.026	-0.586***
	(0.118)	(0.108)	(0.085)	(0.169)	(0.015)	(0.035)	(0.038)	(0.161)
Observations	1,360	1,328	1,321	1,298	1,360	1,328	1,321	1,298
R-squared	0.070	0.198	0.241	0.266	0.065	0.192	0.240	0.266

Annex

Table A1. Average earning for complete secondary education and complete university at regional level, by gender and area of residence

	Complete Secondary education					Complete University					Return to University				
	All	Urban		Rural		All	Urban		Rural		All	Urban		Rural	
		M	F	M	F		M	F	M	F		M	F	M	F
Region															
Amazonas	185	250	75	189	107	317	375	314	300	157	132	125	239	111	49
Ancash	231	331	121	223	150	388	498	330	361	187	156	168	209	138	37
Apurimac	160	251	108	152	115	254	229	259	280	338	95	-22	152	127	224
Arequipa	343	417	242	353	181	386	470	303	382	193	43	52	61	29	12
Ayacucho	182	231	88	222	107	308	373	257	290	208	126	141	169	68	101
Cajamarca	202	295	144	172	85	267	327	252	195	179	65	32	108	23	95
Callao	323	377	253	0	0	419	468	373	0	0	95	91	120	0	0
Cusco	254	303	156	253	220	311	396	248	296	216	57	93	93	43	-4
Huanuco	210	297	180	199	109	289	299	274	248	367	79	2	95	49	258
Ica	285	339	214	262	245	330	375	282	385	332	45	36	68	123	87
Junin	254	356	153	237	139	279	335	242	325	205	25	-21	89	88	66
LaLibertad	242	324	111	361	111	346	393	304	370	350	104	69	193	8	239
Lambayeque	226	299	118	256	98	291	351	241	241	237	66	52	123	-15	140
Lima	328	372	265	295	166	455	529	397	544	243	127	157	132	250	77
Loreto	221	285	160	148	71	331	428	246	0	0	109	143	86	0	0
MadreDeDios	337	438	224	299	357	425	578	375	489	222	89	140	151	190	-135
Moquegua	330	452	152	405	72	451	553	319	435	261	121	101	168	29	188
Pasco	289	361	154	264	117	345	364	352	372	151	56	3	197	108	34
Piura	252	309	146	174	151	310	414	242	313	252	57	106	97	139	101
Puno	267	719	124	217	91	253	281	247	245	142	-15	-438	123	28	51
SanMartin	236	257	219	229	151	365	350	372	456	0	129	93	153	227	0

Tacna	358	393	313	403	243	336	359	312	524	185	-23	-34	0	121	-58
Tumbes	238	264	189	299	113	289	310	275	351	262	51	47	85	52	148
Ucayali	272	318	210	269	126	320	397	250	603	132	47	79	40	334	6

Note: 2010-2016 ENAHO data; sample restricted to 24-25 years old and to Young Lives provinces only. Earnings expressed in USD using 2016 as year base (2016 average PEN/USD exchange rate of PEN 1= USD 0.2965).

Annex A2. Derivation of earnings distribution from subjective expectations data

The distribution of expected future earnings depend on the expected probability to be employed and what expectations the individual have about the earnings she will be able to get in the two educational scenarios, $E(y)_{se}$ for secondary education and $E(y)_u$ for university (conditional on working). As far as *E(y)* is concerned, Young Lives provides information on subjective expectations of the minimum y_m, maximum earnings y_M, and the probability mass p to the right of the mid-point of the support, with
$p = Prob(y \geq (y_m + y_M)/2)$.

Given a certain distributional assumption, we use these three pieces of information to derive the probability density function of earnings. Following Attanasio and Kaufmann (2014), we do assume a triangular functional form. To determine the mode as first step and then compute the probability density function, and the first moment of the cumulative distribution function, i.e. the expected value of earnings we have to distinguish between four cases as follow:

Case 1: $0 < p < 0.25$ $\rightarrow E(y) = \frac{1}{6}(y_M + y_m(5 - 4p) + 4y_M p)$

Case 2: $0.25 < p < 0.5$ $\rightarrow E(y) = \frac{y_m + y_M + m}{3}$, where the mode $m = y_M - \frac{y_M - y_m}{4p}$

Case 3: $0.5 < p < 0.75$ $\rightarrow E(y) = \frac{y_m + y_M + m}{3}$, where the mode $m = y_m - \frac{y_M - y_m}{4(1-p)}$

Case 4: $0.75 < p < 1 \rightarrow E(y) = \frac{1}{6}(y_M + y_m(5 - 4p) + 4y_M p)$

Finally, we compute the second moment of the cumulative distribution function, i.e. the variance, as follow:

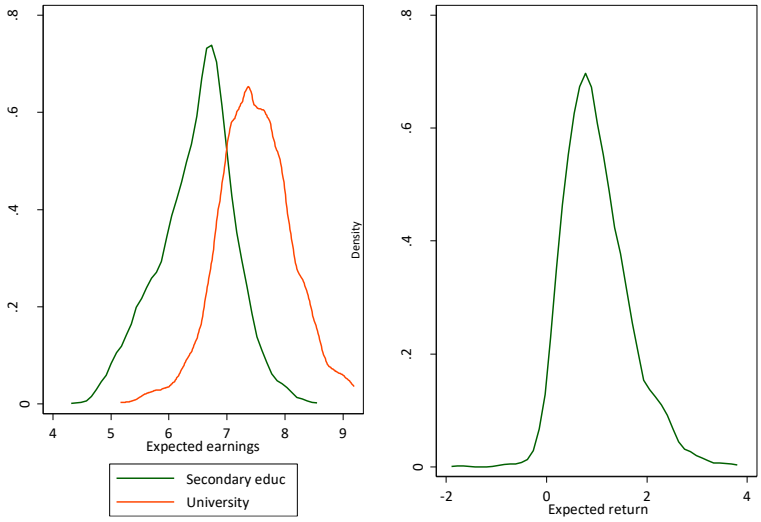
Case 1: $Var(y) = \frac{1}{72}(y_m - y_M)^2(1 + 20p - 32p^2)$

Case 2 and 3: $Var(y) = \frac{1}{18}(y_m^2 + y_M^2 + m^2 - y_m y_M - y_m m - y_M m)$

Case 4: $Var(y) = \frac{1}{72}(y_m - y_M)^2(-11 + 44p - 32p^2)$

The variance is then used to compute the standard deviation of the distribution function as a measure of the perceived earning risk. Furthermore, the return to university is computed as the difference between the expected earnings associated to secondary education and university. The analysis is performed in terms of log earnings. Finally, prior to the questions on earnings expectations conditional on being employed, individuals are asked about the likelihood that they will be employed at the age of 25 for the two different schooling scenarios, secondary education and university.

Figure A1. Distribution of expected earnings and return to tertiary education



Note: Young Lives data, Younger Cohort, Round 5 (2016).

Table A3. Earnings expectations vs. earnings realizations for secondary education ate regional level

Region	Earnings expectations					Expectations - realizations				
	All	Urban		Rural		All	Urban		Rural	
		M	F	M	F		M	F	M	F
Amazonas	252	260	237	396	182	68	10	163	207	75
Ancash	207	220	216	188	194	-24	-110	95	-35	44
Apurimac	253	308	156	279	227	93	57	48	127	112
Arequipa	257	322	201		208	-86	-95	-41		26
Ayacucho	227	291	206	254	171	45	59	117	32	64
Cajamarca	233	256	209	89	198	31	-38	65	-83	114
Callao	136	152	104			-187	-224	-149		
Cusco	242	267		169	267	-11	-36		-84	47
Huanuco	234	293	218	236	190	24	-4	39	38	81
Ica	208	225	185			-77	-114	-28		
Junin	259	313	257	265	228	5	-43	104	29	89
LaLibertad	230	257	190			-12	-67	79		
Lambayeque	142	156	139			-83	-144	21		
Lima	249	263	234	302	274	-79	-109	-31	8	108
Loreto	170	133		208		-51	-152		59	
MadreDeDios	326			326		-10			27	
Moquegua	667	667				337	216			
Pasco	226	44	119		371	-63	-316	-36		254
Piura	223	241	190	273	186	-30	-68	44	99	36
Puno	250	283	226	104	222	-18	-436	102	-114	131
SanMartin	217	251	167	338	205	-19	-6	-51	109	55
Tacna	280	237	294			-79	-156	-18		
Tumbes	240	285	195	293		2	21	5	-6	
Ucayali	482		445	519		209		235	250	
Average	259	260	209	265	223	-1	-80	38	41	88

Note: 2010-2016 ENAHO data; sample restricted to 24-25 years old and to Young Lives provinces only. Earnings expressed in USD using 2016 as year base (2016 average PEN/USD exchange rate of PEN 1= USD 0.2965).

Table A4. Earnings expectations vs. earnings realizations for university at regional level

Region	Earnings expectations					Earnings expectations - realizations				
	All	Urban		Rural		All	Urban		Rural	
		Male	Female	Male	Female		Male	Female	Male	Female
Amazonas	667	758	657	554	584	350	383	343	254	428
Ancash	584	677	653	447	487	196	178	323	86	300
Apurimac	668	815	511	738	584	414	586	252	459	246
Arequipa	712	740	689		637	326	270	386		444
Ayacucho	630	878	543	596	500	322	506	286	306	291
Cajamarca	668	738	598	519	480	401	411	346	323	301
Callao	791	1,001	371			372	533	-2		
Cusco	760	830		341	1,038	449	435		45	822
Huanuco	624	721	495	766	473	335	422	221	519	106
Ica	657	583	762			327	208	480		
Junin	672	678	558	711	674	393	343	316	386	469
LaLibertad	647	717	542			301	325	238		
Lambayeque	273	215	287			-19	-136	46		
Lima	674	730	619	593	445	219	201	222	49	202
Loreto	437	208		667		106	-220			
MadreDeDios	712			712		286			222	
Moquegua	1,186	1,186				735	633			
Pasco	667	815	667		593	322	452	315		442
Piura	539	599	465	672	395	229	185	223	359	143
Puno	681	778	616	222	445	428	497	369	-23	303
SanMartin	559	629	516	621	375	194	279	144	165	
Tacna	482	356	524			146	-3	212		
Tumbes	604	685	474	1,557		315	375	199	1,206	
Ucayali	941		549	1,334		622	0	299	731	
Average	660	697	555	691	551	324	298	261	339	346

Note: Earnings realizations data: 2010-2016 ENAHO data; sample restricted to 24-25 years old and to Young Lives provinces only, population weights used. Earnings expectations: Young Lives data, Younger Cohort, Round 5 (2016). Earnings expressed in USD using 2016 as year base (2016 average PEN/USD exchange rate of PEN 1= USD 0.2965). The return to university is computed as the difference between the (realized) earnings of those who completed a university degree and those who completed secondary education.

Table A5. Average expected earnings, earning risk and probability to find a job

	University	Std. Dev	Secondary	Std. Dev	p-value
Probability of working	0.85	(0.005)	0.64	(0.007)	0.000
Expected monthly earning (PEN)					
(Average) min earnings	479	(8.340)	177	(2.875)	0.000
(Average) max earnings	783	(13.974)	295	(5.234)	0.000
(Average) mid earnings	631	(10.713)	236	(3.848)	0.000
Probability to earn at least mid-monthly earnings	0.75	(0.005)	0.62	(0.006)	0.000
(Average) return to tertiary	395	(9.705)			
Expected earnings (log)	7.49	(0.016)	6.48	(0.016)	0.000
Earnings risk (std)	0.10	(0.002)	0.12	(0.003)	0.000

Note: Young Lives data, Younger cohort, Round 5 (2016). Earnings expressed in USD using 2016 as year base (2016 average PEN/USD exchange rate of PEN 1= USD 0.2965).