# DOCUMENTO DE DISCUSIÓN

DD/11/14

# "Higher Education Decisions in Peru: On the Role of Financial Constraints, Skills, and Family Background"

*Juan F. Castro, Gustavo Yamada y Omar Arias*

UNIVERSIDAD DEL PACÍFICO
CENTRO DE INVESTIGACIÓN

 DD/11/14

**Documento de Discusión**

# "Higher Education Decisions in Peru: On the Role of Financial Constraints, Skills, and Family Background"

Juan F. Castro, Gustavo Yamada y Omar Arias

December 2011

**Abstract**

This paper analyzes the relative importance of short term financial constraints vis a vis skills and other background factors affecting schooling decisions when explaining access to higher education in Peru. We focus on college access disparities between rich and poor households. We use a novel household survey that includes special tests to measure cognitive and non-cognitive skills of the urban population age 14-50. These are complemented with retrospective data on basic education and family socioeconomic conditions in a multinomial model. We find that strong correlation between college enrollment and family income in urban Peru is not only driven by credit constraints, but also by poor college readiness in terms of cognitive skills and by poor family and educational backgrounds affecting preferences for schooling. Family income explains, at most, half of the college access gap between poor and non-poor households. The other half is related to differences in parental education, educational background and cognitive skills. Our results indicate that credit and/or scholarship schemes alone will not suffice to change the regressive nature of higher education enrollment in Peru, and that such programs will face strong equity-efficiency trade-offs.

Keywords: Higher education, cognitive skills, non-cognitive skills, credit constraints, Peru.

Correo de los autores: castro_jfl@up.edu.pe, yamada_ga@up.edu.pe y oarias@worldbank.org

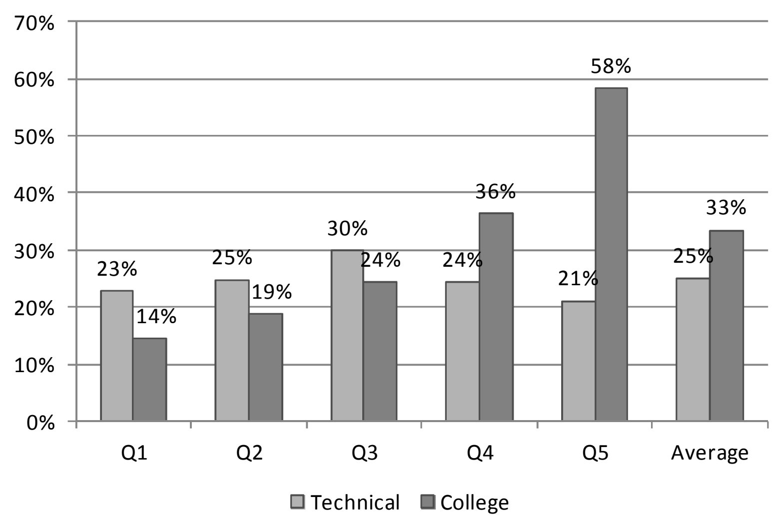
* Authors would like to thank Roberto Asmat, Fernando Mendo and David Vera-Tudela for their skillful research assistance, and are thankful for useful insights and comments by participants at seminars at the World Bank in presentations of earlier versions of the paper. Juan Francisco Castro and Gustavo Yamada acknowledge financial support from the World Bank. The opinions expressed herein do not represent the views of the World Bank, its Executive Directors, or the governments they represent. All remaining errors are the authors' responsibility.

# 1. Introduction and motivation

Access to higher education in Peru is remarkably regressive. According to the latest national household survey (ENAHO 2010), in the bottom 20% of the income distribution, only 37% of individuals with completed secondary were able to enroll in some type of higher education. In contrast, nearly 80% of youngsters in the richest 20% had access to this educational level (see Figure 1).

Interestingly, this average pattern is driven by the regressive nature of college (university) access. Access to non-university (technical) higher education does not exhibit the same regressive pattern. Demand for this type of education is a non-monotonic function of family income and, as individuals become richer, they tend to move away from this type of degree in favor of a university degree.

**Figure 1**
17-25 year olds with completed secondary education that had access to higher education
(% of the urban population)




Source: ENAHO (2010).

Until recently, the empirical economics literature on the determinants of access to post-secondary education had focused largely on the influence of family socio-economic conditions. This literature has often emphasized that family income and/or the socioeconomic status of parents, a proxy for credit constraints, have a strong influence on the probability of attending college. Thus, limited access to credit is usually presented as the main reason behind a regressive pattern such as the one portrayed above.

Families are rationed from credit markets for human capital investments (due to information asymmetries regarding their return) and depend on their endowments to finance the direct and opportunity costs of higher education. Poor families fall short of resources and, thus, fail to make this type of investment despite its potentially high return[1]. The implication is that income transfers or college loan programs to the low income population would be an effective way of increasing their rate of college attendance.

The above, however, is only part of the story. The early literature on human capital (Becker, 1964) emphasizes the role of cognitive ability in the determining school achievement and human capital formation. The signaling literature (Spence, 1973) posits education as a signal of a person's overall ability level, equated to cognitive skill. Cameron and Heckman (2001) and Carneiro and Heckman (2002) provide empirical evidence suggesting that cognitive abilities are more binding than credit constraints in determining access to college.

In addition, psychologists and sociologists have long studied how a person's multitude of abilities, behavioral traits and motivation are crucial to understand the pursuit and achievement of long-term goals such as post-secondary careers (e.g, see the reviews in Farkas, 2003, and Eccles and Wigfield, 2003).

More recently, starting with the work of Bowles and Gintis (1976), with the availability of relevant data the economics literature have redressed the study of higher education attainment to consider the role of the so-called "non-cognitive" (soft) skills. Bowles and Gintis (1976) show evidence that perseverance, dependability, and consistency are among the most important predictors of grades in school. See the reviews in Heckman (2000), Carneiro and Heckman (2002), Cunha, *et*

[1] It is well documented for Peru that average returns for university higher education are large (around 17% after factoring in direct costs) and considerably larger than returns for technical higher education (see Yamada and Castro, 2010).

*al.,* (2006) in the labor economics literature, and Barrick, *et al.,* (2001, 2005) and Roberts *et al.,* (2007) for reviews of the psychological literature.

In two salient studies, Heckman, *et al.* (2006) and Heckman and Rubinstein (2001) find that in the United States non-cognitive skills are quantitatively important determinants of post-secondary educational attainment. Controlling for measured cognitive ability, they use data from the U.S General Educational Development (GED) high school equivalency testing program to show that GED recipients fail to complete high school and hence to pursue a college education because they lack in non-cognitive skills such as self-discipline and perseverance.

Much of the limited attention to skills in analyses of post-secondary schooling attainment is due to the lack of reliable data measurement, especially in developing countries. In Latin America, a recent new survey of Chilean youth have measured several cognitive and non-cognitive skills and document their importance in predicting socio-economic outcomes including higher education attainment (Bassi and Galiani, 2010).

In Peru, the World Bank has recently collected the first nationally representative household survey (ENHAB 2010) in the region that includes standardized psychology and achievement tests to measure the cognitive and non-cognitive skills of the working-age population living in urban areas. Skill measures (Rasch and z-scores) are based on cognitive tests of numeracy and problem-solving ability, working memory, verbal fluency and receptive language, and self-reported responses to scales of the Big-Five Personality Factors (Goldberg, 1990) and Grit (perseverance and the will to strive for long term goals; Duckworth *et al.*, 2007). The survey also includes a comprehensive questionnaire to assess individual educational histories (from pre-school through college/technical education), family background and socio-economic conditions. The latter includes data on parental education and occupation, family size and composition, and schooling trajectories related to access (e.g., distance) and characteristics of institutions attended, self-reported scholastic aptitudes, parental involvement, self-reported family economic conditions at the time of attendance to basic education, choice of post-secondary career and institution and reasons.

This data offers a unique opportunity to explore the determinants of higher education decisions in Peru and, in particular, to assess the importance of short term credit constraints when explaining the regressive nature of enrollment. Our objective, thus, is to measure the relative importance of short term financial constraints *vis a vis* skills and other background factors affecting "tastes for education" when explaining access to higher education in Peru, as well the disparities between rich and poor households regarding access to college education. This analysis should help dimension the potential effect of credit and/or scholarship schemes on higher education access among low income families, as well as the potential efficiency-equity trade-offs that such policies entail.

The rest of the paper is organized as follows. Section 2 briefly reviews the concepts of skills and describes the measurements used for our analysis. Section 3 sketches the relationship between family income, skills and higher education enrollment in Peru captured in the ENHAB survey. Section 4 presents our model and empirical strategy, and section 5 discusses the results. Section 6 concludes with some observations on the implications of our findings for public policies to improve access to tertiary education and some directions for future research.

## 2. Skills: concepts and measurement in Peru

The recent labor economics literature distinguishes four types of marketable skills: cognitive (e.g., verbal/literacy, numeracy, problem-solving), non-cognitive[2] (e.g., self-discipline, perseverance, dependability, team work) –also called "soft"—, technical and professional (e.g., vocational, career qualifications) and job-specific acquired through work experience. Due to lack of data, until recently it had not been possible to give an adequate account of these various skills, how they are developed (at homes and schools), and to document their relationship with higher education decisions and reward in labor markets.

[2] Non-cognitive skills are more appropriately referred as socio-emotional skills or social literacy by psychologists as these involve processes of cognition. We stick here to the now conventional –although misleading- jargon used in the economics literature.

Skills formation is a cumulative life-cycle process. It can be thought as climbing a ladder since very early in life: as individuals age they build on the learning in each step to move up to the next step. There is a large body of literature documenting the importance of adequate health and nutrition during the so called "first 1,000 days" – from conception throughout the first 2+ infant years— in the development of basic cognitive and socio-emotional abilities and readiness to learn at school and in the adult life (Shonkoff and Phillips, 2000; Heckman and Cunha 2010). The quality of nurturing environments during infancy and childhood further develop cognitive ability and also shapes socio-emotional traits.

There are different sensitive periods for the formation of these multiple skills, where both heritability and environmental influence play a role[3]. While basic cognitive ability is well set by the teen years, formal schooling provide with subject knowledge and tools that enhance the cognitive capacity to undertake tasks and solve new problems. Socio-emotional skills continue to develop and remain malleable through the adolescence and early adult years (Cunha and Heckman, 2010; Heckman, 1996, 2004). These skills determine a person's "readiness to learn" over the life cycle by shaping the capacity and motivation to absorb new knowledge, adapt and solve new problems. Thus, these skills correlate with higher educational attainment. In particular, professional and technical skills are developed through tertiary schooling and training (formal or on-the-job), and job-specific skills then acquired through labor market experience.

Much of the literature on cognitive tests argues that one dominant factor ("G") can summarize a person's cognitive ability and performance in cognitive tests. This has not been the case in the personality psychology or social literacy literature given the multitude of distinct behavioral traits subsumed under the category of non-cognitive skills.

The National Skills and Labor Market Survey (ENHAB) in Peru was designed over 1 year and the data collected during Jan-March 2010 as a self-standing nationally

[3] A solid body of evidence from biology (epigenetics), neuroscience, psychology, and education supports a consensus that the "Nature" vs "Nurture" distinction is obsolete and vindicates the power of public intervention to influence cognitive and socio-emotional abilities (Shonkoff and Phillips, 2000; Cunha and Heckman, 2010).

representative household survey covering urban areas (2,666 households from cities with population >70,000), the Coast, Highland, Jungle, and Metropolitan Lima. The survey instrument uses the same modules of Peru's regular household survey for housing living conditions, demographics, educational attainment, employment/income (almost identical), and supplement these with modules to collect new data on cognitive skills and personality traits applied to a random sub-sample of the population age 14-50.

The cognitive tests include the PPVT-4 (a widely used standardized test of receptive language), and a battery of tests specifically designed to measure verbal ability, working memory, and numeracy/problem-solving. Socio-emotional skills are captured with self-reported tests for personality traits related to behaviors which the labor economics and psychology literatures suggest are important for labor market outcomes. The latter are measured with scales of the Big-five Personality Factors (Openness to experience; Conscientiousness; Extraversion; Agreeableness; Emotional Stability), widely accepted in psychology to characterize differences in broad personality traits (and associated behaviors), and Grit, a narrower trait capturing one's inclination and motivation to achieve long term goals through perseverance of effort and consistency of interest (Duckworth, *et al.* 2007). Appendix 1 contains a more detailed description of the tests. For details on the methodology for constructing the tests and the resulting test scores see Cueto, *et al.* (2010) and Claux and La Rosa (2010).

The survey also included retrospective questions on school trajectories related to access (e.g., distance), characteristics of institutions attended as school quality proxies, self-reported scholastic aptitudes and performance, parental involvement, family economic conditions, choice of post-secondary career and institution and reasons for choice. It also collects unusually detailed data on family background, including parental (father and mother) education and occupation, family size and relation to siblings (number, gender, birth order), place of birth and residence. In what follows we provide a description of the most relevant variables for the questions addressed in our analyses.

## 3. Socio-economic conditions, skills and access to higher education in Peru

Unlike other countries of the region (like Colombia or Chile) and almost every developed country, Peru lacks a publicly subsidized credit scheme for higher education investments. While this advocates for the role of short term financial constraints when explaining the regressive nature of enrollment documented above, we should also notice that Peru has direct public intervention in the higher education market. Public universities and institutes are tuition free and currently concentrate nearly 40% of higher education enrollment. Surprisingly, this intervention is far from alleviating the regressive nature of higher education access. As documented in Morón *et al.* (2009), two thirds of students enrolled in a public institute and 80% of those enrolled in a public university come from non-poor households.

The regressive nature of the tuition-free public supply is a first piece of evidence pointing towards the fact that other constraints (besides short term financial constraints) must be in place preventing poor households from investing in college education. If we further inquire inside our national household survey about the reasons behind the failure to enroll in higher education, nearly 40% of respondents invoke monetary reasons but also a significant 28% argue that they have already finished their education and/or they are not interested in pursuing a college degree.

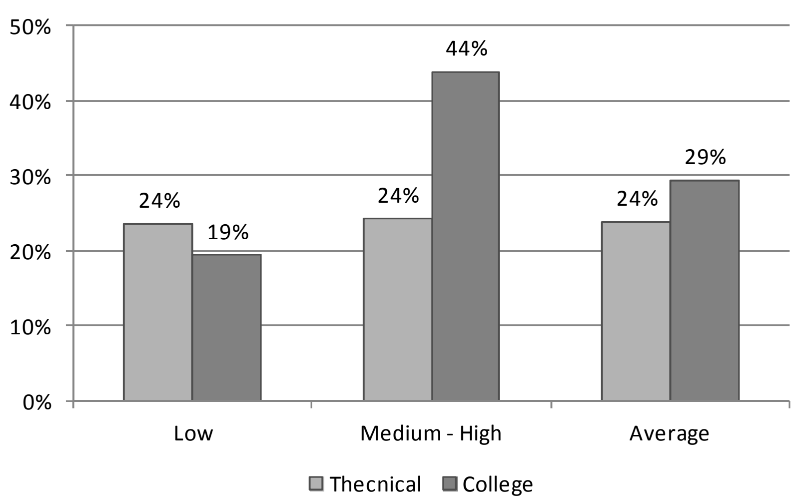
This last result is clearly at odds with current estimates of the average return to higher education in Peru, which can be as high as 17% even after factoring in direct costs (see Yamada and Castro, 2010). Information asymmetries regarding this high return can be part of the explanation but it also reasonable to explore other barriers to entry, the existence of other costs besides pecuniary, and the possibility that average returns could be masking important differences between expected returns for individuals from different family backgrounds.

Basic skills developed early in life are related to individuals' success in selection processes to access higher education, the extent of psychic costs related to the study process and, eventually, the probability of graduation and their degree of certainty regarding the promise of increased earnings. Family and educational background, on the other hand, can also affect individuals' preferences or "tastes for education" and, thus, the acquisition more schooling after completing secondary education.

Tight correlation between family income and higher education enrollment can be due to the existence of binding financial constraints. However, it can also be provoked by the fact that poor family environments and low basic education quality affecting the skill formation process and preferences for schooling are also correlated with current income.

Information contained in the ENHAB survey not only confirms the regressive nature of college access but also reveals a significant correlation between family income and individuals' basic skills, especially cognitive skills. In a sample of working age individuals, current family income is a biased indicator of the availability of resources at the time postsecondary choices were made. In the adult population, current family income is caused by postsecondary choices. Thus, we use respondents' self-report of the socioeconomic status of their families while he attended secondary school as a proxy.

**Figure 2**
Individuals with completed secondary
education that had access to higher education
(% of the urban population)

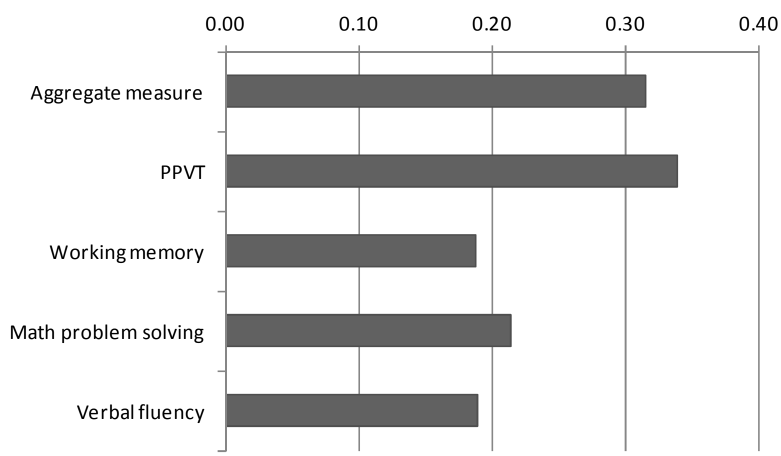



Source: ENHAB (2010).

Perceived socioeconomic status during secondary education in ENHAB is expressed through three possible categories: low, medium, or high. Results shown in Figure 2 confirm that family socioeconomic status has a strong correlation with college access. Moving from a low to a medium-high socioeconomic status[4] raises college access in more than 24 percentage points and is comparable to moving from the second income quintile to some point between the fourth and fifth quintile (see Figure 1).

Figures 3 and 4 compare cognitive and non-cognitive test scores for the same two socioeconomic groups. Results reveal important gaps especially in the cognitive domain: all cognitive measures exhibit statistically significant differences and, thus, also the aggregate measure used in the analysis that follows.

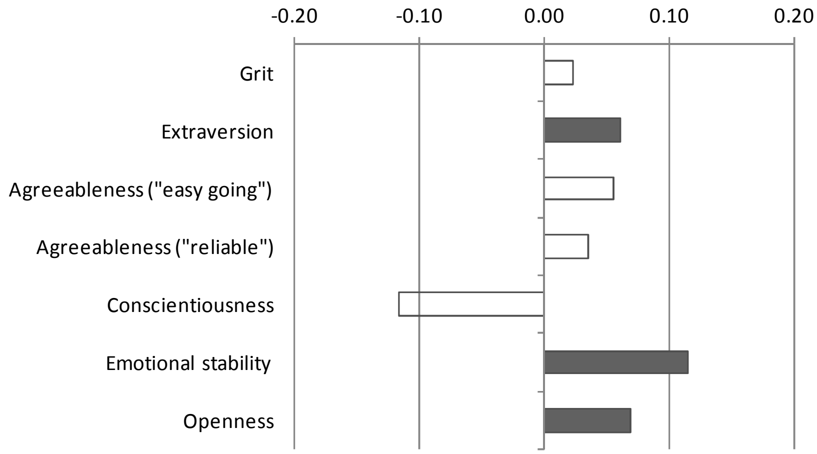
**Figure 3**
Differences in cognitive skills: low vs. medium-high socioeconomic status individuals
(in standard deviations)




*Shaded bar denotes statistically significant gap at 5%.

[4] We decided to group these last two categories because only 0.63% of the sample reported a high socioeconomic status. Proportions of each category in the sample are: 59.2% low and 40.8% medium-high.

**Figure 4**
Differences in non-cognitive skills: low vs. medium-high socioeconomic status individuals
(in standard deviations)




*Shaded bar denotes statistically significant gap at 5%.

## 4. Model and empirical strategy

### 4.1 Why should skills and family background matter?

Our analysis focuses on postsecondary school trajectories and, in particular, on the decision of pursuing higher education and the type of higher education chosen (college vs. technical).

As already discussed, monetary resources are relevant when explaining higher education decisions in the presence of credit constraints. From a broader perspective, Checchi (2006) argues that educational choices exhibit intergenerational persistence because these choices are conditioned by abilities, financial resources, and family and cultural background, all of which also exhibit some degree of persistence.

The role of ability can be motivated using a simple static schooling decisions model as the one proposed in Card (1994) or in Card and Krueger (1996). We can assume individual $i$ chooses schooling to maximize utility given by:

$$U(y_i, S_i) = log(y_i) - f(S_i) \quad (1)$$

Where $y_i$ represents earnings, $S_i$ are years of schooling, and $f(\cdot)$ is a convex cost function. The first order condition satisfied by the optimal number of schooling years $(S_i *)$ is given by:

$$\frac{\partial log(y_i)}{\partial S_i} = f'(S_i) \tag{2}$$

The above simply states that the individual will choose the number of schooling years that equates the marginal cost with the marginal revenue of schooling. The equation relating log-earnings to schooling is a crucial piece of the model (it represents the budget constraint faced by the individual) and we can assume it has a person specific slope (as in Card, 1994) or a person specific intercept (as in Card and Krueger, 1996) capturing individual "ability".

In the first case, ability or skills directly determine the marginal return to schooling and will directly affect the optimal number of schooling years chosen: individual skills are an argument of $\frac{\partial log(y_i)}{\partial S_i}$. Thus, more skills will directly lead to more schooling because they raise its return.

In the second case, the positive relation between skills and the optimal number of schooling years will depend on the correlation between skills and the marginal cost of schooling $(f'(S_i))$. It should be noticed that, in the above formulation, the marginal cost of schooling captures the marginal rate of substitution between schooling and future earnings. It, therefore, has an individual specific component that accounts for differences in access to funds (short term financial constraints) but also for differences in tastes and aptitudes for schooling.

The above is related to the role of family and cultural background influences on schooling decisions discussed in Checchi (2006). Checchi argues that children of educated parents are more likely to acquire education through imitation and induced educational choices: educated parents are aware of the social and economic value of education and will put more pressure on their children to acquire education. More educated parents are also in advantage regarding access and use of information regarding school quality and can orient their children towards better opportunities[5].

An important implication of the above is that controlling only for skills will not suffice to approximate the relative importance of financial constraints. While we can suspect a negative correlation between skills and schooling costs (Card (1994) and Card and Krueger (1996)), these costs are determined by a very broad set of family and cultural influences. While it is not our intention to disentangle and discuss each of these effects, we certainly do have to control for a comprehensive set of background variables to have a reliable estimate of the effect of financial resources.

### 4.2 Empirical specification

Given constraints, the agent will choose the postsecondary trajectory that maximizes her net utility. With a total of $J$ trajectories, we assume agent's $i$ net utility from choosing trajectory $j$ can be expressed as follows:

$$u_{ij} *= x_i'\beta_j + \varepsilon_{ij} \quad (1)$$

Where $x_i$ is a vector of observed characteristics of agent $i$, and $\varepsilon_{ij}$ is an idiosyncratic error term uncorrelated with $x_i$. Elements in vector $x_i$ in this reduced form equation should affect the feasible set of higher education choices available to the household as well as perceived costs and benefits of these choices.

The formulation discussed above constitutes the basis for an unordered multinomial discrete response model. Net utility derived by agents is not observable. Instead, we observe their choices measured in the form of a discrete variable or category $(y_i = 1, 2, \dots J)$.

[5] This latter effect can be particularly important when the educational system is not homogeneous like in Peru.

Given an assumption for the distribution of the error term in (1), our basic formulation allows us to model the probability of choosing a particular alternative $(k)$. In particular, agent chooses alternative $k \in \{1, \dots, J\}$ such that:

$$k = argmax\,\{u_{ik}*\} \tag{2}$$

This, in turn, implies that:

$$\begin{aligned} \Pr(y_i = k) &= \Pr\big(u_{ik}* > u_{ij}*\big)\ \forall\, j \neq k \\ &= \Pr\big(x_i'\beta_k + \varepsilon_{ik} > x_i'\beta_j + \varepsilon_{ij}\big)\ \forall\, j \neq k \\ &= \Pr\big(\varepsilon_{ik} - \varepsilon_{ij} > x_i'\beta_j - x_i'\beta_k\big)\ \forall\, j \neq k \end{aligned} \tag{3}$$

If we allow error terms to follow a logistic distribution, the above yields the multinomial logit model, which is widely used to model choice between more than two alternatives based on individual characteristics. In particular, the logistic distribution assumption and (3) imply that:

$$\Pr(y_i = k) = \frac{\exp\,(x_i'\beta_k)}{\sum_{j=1}^{J} \exp\,(x_j'\beta_k)} \tag{4}$$

Based on (4) we can fully characterize a likelihood function to obtain ML estimates of parameter vectors for each category. These, in turn, can be used to estimate the marginal effect of a covariate on the probability of choosing a certain category, or to predict this probability for an individual with certain characteristics.

Our multinomial model comprises three categories: (i) did not enroll in higher education; (ii) did enroll in technical higher education; (iii) did enroll in university higher education. These categories were built using the entire sample of ENHAB respondents with completed secondary education.

Following our objective, the variables we include in vector $x_i$ are respondents' self-report of the socioeconomic status of their families while he attended secondary school, measures of cognitive and non-cognitive skills, and a comprehensive set of family and education background controls.

Covariates are grouped in five categories: (i) socioeconomic status; (ii) cognitive skills; (iii) non-cognitive skills; (iv) parental background; and (v) educational background. Socioeconomic status was introduced through a dummy variable indicating if the individual reported a medium or high condition. Parental background variables include parents' educational attainment and respondents' perception regarding the importance given by their parents to their education. Educational background controls include access to preschool education, if the school attended was public, and variables reflecting individuals' performance and effort at school.

### 4.3 Model simulation

In addition to the estimation and discussing of marginal effects for specific variables within our five groups of covariates, we are particularly interested in measuring their relative importance when explaining average enrollment in higher education as well as the regressive nature of access to college education.

To accomplish this we need a benchmark for the variations induced in our covariates since the effects of marginal or unit changes are not directly comparable. The benchmarks we propose are: (i) the differences between an average individual and those who had access to higher education; and (ii) the differences between individuals who report a low socioeconomic status and those who report a medium or high condition.

With (i), the objective is to measure the contribution of each covariate difference in closing the gap between observed and full (100%) access. With (ii), the idea is to measure the contribution of each covariate difference in closing the gap between college access rates for low socioeconomic status and medium-high socioeconomic status individuals. We seek to measure the relative importance of these differences when we account for the regressive pattern of college enrollment.

Let us define values for $y_i$ as 0, 1 and 2 for categories: failed to enroll in higher education, enrolled in technical education, and enrolled in college education, respectively. Recall that variables in the vector of covariates $x_i$ were classified in five groups. Allow vector $x_{ig}$ for $g = 1, \ldots, 5$ contain the covariates in each group, while $x_{i\tilde{g}}$ contain the rest of covariates such that $x_i{}' = [x_{ig}{}' \quad x_{i\tilde{g}}{}']$. Finally, allow vector $\bar{x}$ contain covariate values for an average individual, and vectors $\bar{x}_E$, $\bar{x}_L$, and $\bar{x}_H$ contain mean covariate values for those who had access to higher education, report a low socioeconomic status, and report a medium or high condition, respectively.

With this, the contribution of differences in covariate group $g$ when closing the gap between observed and full higher education enrollment $\left(\%G1_g\right)$ can be expressed as:

$$\%\mathrm{G1_g} = \frac{\left[1 - \Pr\left(y_i = 0|x_{ig} = \bar{x}_{Eg}, x_{i\tilde{g}} = \bar{x}_{\tilde{g}}\right)\right] - \left[1 - \Pr(y_i = 0|x_i = \bar{x})\right]}{\left[1 - \Pr(y_i = 0|x_i = \bar{x}_E)\right] - \left[1 - \Pr(y_i = 0|x_i = \bar{x})\right]}$$

$$= \frac{\Pr(y_i = 0|x_i = \bar{x}) - \Pr\left(y_i = 0|x_{ig} = \bar{x}_{Eg}, x_{i\tilde{g}} = \bar{x}_{\tilde{g}}\right)}{\Pr(y_i = 0|x_i = \bar{x}) - \Pr(y_i = 0|x_i = \bar{x}_E)} \tag{5}$$

In the expression above, $\Pr\left(y_i = 0|x_{ig} = \bar{x}_{Eg}, x_{i\tilde{g}} = \bar{x}_{\tilde{g}}\right)$ refers to the probability of failing to enroll in higher education of an average individual with covariates of group $g$ evaluated in the mean value of those who had access to higher education.

In a similar way, we can calculate the contribution of differences in covariate group $g$ when closing the gap between college access rates for low socioeconomic status and medium-high socioeconomic status individuals $\left(\%G2_g\right)$ following:

$$\%\mathrm{G2_g} = \frac{\Pr\left(y_i = 2|x_{ig} = \bar{x}_{Hg}, x_{i\tilde{g}} = \bar{x}_{L\tilde{g}}\right) - \Pr(y_i = 2|x_i = \bar{x}_L)}{\Pr(y_i = 2|x_i = \bar{x}_H) - \Pr(y_i = 2|x_i = \bar{x}_L)} \tag{6}$$

### 4.4 Endogeneity issues

There are two potential sources of endogeneity in our empirical analysis. The first one has to do with the fact that our sample comprises secondary school graduates and that the discrimination processes involved in basic education makes a case for a non-random sample.

Omitted cognitive and non-cognitive skills are usually the source of selection bias when we work with a subgroup of individuals that have completed a certain schooling level (and especially non-cognitive skills[6]). We are now able to control directly for these skills, and could claim that there should be no source of correlation between the error terms of a selection equation (that models the probability of concluding high school), and our main equation that models college enrollment. Despite this, we built a bivariate probit model to test for selection bias arising from individuals who drop out before completing high school. We simultaneously model secondary education completion and higher education enrollment allowing for correlation between the error terms of both equations. Results (presented in Appendix 2) show no evidence of selection bias in the sample of secondary school graduates[7].

Another source of potential bias in our empirical strategy has to do with the possibility of reverse causality between college attendance and test scores. A significant body of empirical literature supports the fact that cognitive skills are developed early in life (around age 8) while non-cognitive skills remain more malleable through adolescent years (see Cunha, *et al.* (2006) for a summary of empirical evidence on life cycle skill formation). While this should help us claim that skills are exogenous in a college attendance equation, it should be noticed that our database presents us with measured skills rather than with latent skills. This distinction is particularly important since several studies (see, for example, Hansen, *et*

[6] This is explanation provided by Heckman and Rubinstein (2001) to the findings of wage differentials between high school graduates, GED recipients and drop-outs despite the first two having ‘equivalent’ credentials and measured cognitive abilities.

[7] Absence of correlation between error terms confirms that there is no risk of selection bias arising from high school drop outs. Coefficient signs and significance in the bivariate probit model for the probability of higher education enrolment give high school completion are consistent with multinomial model estimates.

*al.* (2004) and Heckman, *et al.* (2006)) document how measured skills can be affected by late schooling.

As shown in Appendix 3, our particular setting implies that the use of test scores to account for skills introduces two potential biases with opposite effects: (i) an attenuation bias due to measurement error; and (ii) a positive bias due to causality from schooling to test scores. In principle, the latter could be eliminated by working with an adjusted version of test scores: one in which the effect of schooling on measured ability has been washed out. This strategy faces its own problems due to the difficulty to isolate variation in schooling uncorrelated with skills[8].

The above implies a large risk of underestimating the effect of skills on higher education enrollment[9]. On one hand, there is the risk of over-adjusting test scores for those who attended higher education due to failure to find an instrument for schooling that does not correlate with skills. On the other hand, and even if we could successfully purge the effect of schooling on measured ability, the attenuation bias would still remain. For these reasons, we decided to present and discuss the results obtained with the original test scores and leave further exploration of the combined effects of and attenuation and positive bias as an opportunity for future research.

## 5. Results

In the multinomial model, individual coefficient estimates and their significance can be informative of the impact of variables on the probability of choosing an alternative relative to the baseline category. Their interpretation, however, is more complicated when we work with several categories (the absolute effect of a covariate on a certain

[8] Instrumental variable estimation could be applied to search for a consistent estimate of the effect of schooling on measured skills. This estimate could then be used to remove the effect of schooling and work with a "residualized" version of test scores. The success of this strategy heavily depends on the choice of instrument: it should correlate with schooling but should not correlate with the unobservable component of skills. This is difficult to accomplish because skills are the result of a cumulative process and skill formation equations usually lack information related to the effect of early home environments. These affect skills and are correlated with schooling.

[9] We tried several versions of an adjusted test score supposedly purged from the effect of higher education attendance. Instruments used to search for a consistent estimate of the effect of schooling on test scores included distance to school and age, the latter based on the notion that if age affects measured skills in sample of high school graduates it should only be through granting more opportunities to attend higher education. In both cases results obtained suggested we were underestimating the effect of skills: estimated marginal effects were not significant or even negative.

alternative can even have the opposite sign of its coefficient). For this reason, we present coefficient values and their significance in Appendix 2 and focus this discussion on marginal effects and the results of the simulation discussed above.

### 5.1 Marginal effects and their significance

Table 1 shows marginal effects for each covariate on an average individual in four versions of the empirical model. We start with a naive version in which we only include individuals' socioeconomic status (panel A), and progressively control for skills (panel B), parental background (panel C), and educational background (panel D).

Several results are worth highlighting. In the simplest version of the model, a change to medium-high socioeconomic status boosts college access (and reduces the probability of failing to enroll in higher education) in approximately 23 percentage points. This result is consistent with the 26 percent gap shown in Figure 1[10]. As suspected in the motivation of this paper, it would be misleading to attribute the full measure of this effect to the presence of short term financial constraints. As shown in panels B, C, and D, a significant part of this difference is related to family and educational background variables and long term constraints affecting the acquisition of basic skills. In particular, controlling for skills and background variables reduces the impact of a change to medium-high socioeconomic status on college access down to nearly 12 percentage points.

This does not imply, however, that short term financial constraints are not binding at all. Socioeconomic status retains statistical significance even in the more complete version of the model. In fact, is in this version of the model (where we control for skills and all background variables) "socioeconomic status" is a better proxy of monetary resources available to the family by the time postsecondary schooling choices were made.

[10] Note that in panel (A) we already control for individual characteristics such as age, sex, and language.

## Table 1: Marginal effects from the multinomial model

| | (A) | | | (B) | | | (C) | | | (D) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Covariate groups** | Did not enroll | Technical | College | Did not enroll | Technical | College | Did not enroll | Technical | College | Did not enroll | Technical | College |
| ***Socioeconomic status*** | | | | | | | | | | | | |
| Medium or high = 1 | -0.236*** | 0.02 | 0.22*** | -0.224*** | 0.02 | 0.206*** | -0.182*** | 0.02 | 0.158*** | -0.146*** | 0.03 | 0.118*** |
| ***Cognitive skills*** | | | | | | | | | | | | |
| Aggregate measure | - | - | - | -0.228*** | 0.043** | 0.186*** | -0.21*** | 0.043* | 0.167*** | -0.193*** | 0.046** | 0.147*** |
| ***Non-cognitive skills*** | | | | | | | | | | | | |
| Grit | - | - | - | -0.06** | -0.01 | 0.072*** | -0.063*** | -0.01 | 0.073*** | -0.054** | -0.02 | 0.072*** |
| Extraversion | - | - | - | 0.02 | 0.02 | -0.035° | 0.02 | 0.02 | -0.041* | 0.03 | 0.02 | -0.043* |
| Agreeableness ("easy going") | - | - | - | -0.03 | 0.02 | 0.01 | -0.03 | 0.02 | 0.01 | -0.03 | 0.02 | 0.00 |
| Agreeableness ("reliable") | - | - | - | 0.04 | -0.044** | 0.01 | 0.04 | -0.047** | 0.01 | 0.037° | -0.046** | 0.01 |
| Conscientiousness | - | - | - | 0.02 | 0.02 | -0.04* | 0.02 | 0.02 | -0.035* | 0.01 | 0.01 | -0.02 |
| Emotional stability | - | - | - | -0.01 | 0.01 | 0.00 | -0.01 | 0.02 | 0.00 | -0.01 | 0.01 | 0.00 |
| Openness | - | - | - | -0.02 | 0.02 | 0.00 | -0.01 | 0.02 | 0.00 | -0.02 | 0.03 | -0.01 |
| ***Parental background*** | | | | | | | | | | | | |
| Father educational attainment (secondary = 1) | - | - | - | - | - | - | -0.152*** | 0.084* | 0.07 | -0.148*** | 0.084* | 0.06 |
| Father educational attainment (higher = 1) | - | - | - | - | - | - | -0.236*** | 0.02 | 0.215*** | -0.239*** | 0.03 | 0.208*** |
| Mother educational attainment (secondary = 1) | - | - | - | - | - | - | 0.06 | -0.03 | -0.03 | 0.05 | -0.02 | -0.04 |
| Mother educational attainment (higher = 1) | - | - | - | - | - | - | -0.05 | -0.02 | 0.06 | -0.02 | 0.00 | 0.01 |
| Importance given by parents to education (high = 1) | - | - | - | - | - | - | -0.01 | 0.01 | -0.01 | 0.02 | 0.02 | -0.04 |
| Importance given by mother to education (high = 1) | - | - | - | - | - | - | -0.06 | -0.127** | 0.187** | -0.04 | -0.133*** | 0.175** |
| ***Educational background*** | | | | | | | | | | | | |
| Preschool (public = 1) | - | - | - | - | - | - | - | - | - | 0.112** | -0.07* | -0.04 |
| Preschool (private = 1) | - | - | - | - | - | - | - | - | - | 0.10 | 0.01 | -0.112** |
| Public school = 1 | - | - | - | - | - | - | - | - | - | 0.141* | 0.1* | -0.241*** |
| Had to repeat a year or more in school = 1 | - | - | - | - | - | - | - | - | - | 0.121*** | -0.01 | -0.112*** |
| Perception regarding performance (top student = 1) | - | - | - | - | - | - | - | - | - | -0.171** | 0.01 | 0.158** |
| Perception regarding effort (large = 1) | - | - | - | - | - | - | - | - | - | -0.122*** | 0.07* | 0.05 |

Significant at: 1% (***), 5% (**), 10% (*), 15% (°)

All models control for age, sex, first language, birth order, number of siblings, and born in Lima (capital city).

In the full version of the model (panel D), all covariate groups exhibit variables with a significant effect on college access and higher education enrollment. Within the group of skills, our aggregate measure of cognitive ability exhibits the largest effect: nearly 15 percentage points on college access and 20 percentage points on higher education access, for an increase of one standard deviation. Grit is also significant but with a considerably smaller effect. Goldberg's "big five" personality traits fail to show significant results except for "extraversion" and "agreeableness" which appear to work against enrollment in college and technical higher education, respectively.

It is interesting that variables within the parental and educational background groups have a significant contribution even after controlling for skills. In fact, omission of these variables would have led to a biased assessment of the effect of short term monetary constraints: the marginal effect of "socioeconomic status" declines considerably after their inclusion. Within the parental background group, presence of a father with higher education has an important effect on enrollment. Besides determining resources available during childhood that affect skill formation, fathers' educational attainment must also be reflecting household "tastes for education". Interestingly, maternal concerns regarding individuals' school performance are also related to more college enrollment against enrollment in technical education.

It is also interesting that variables reflecting individuals' educational background (such as type of school attended: public or private) also remain significant. Peru's basic education system exhibits a significant quality gap against public schools. In that sense, the effect of the variable indicating public school attendance should not be surprising. The fact that this variable retains significance and a sizeable effect after controlling for basic cognitive and non-cognitive skills, however, reflects that school environments can also affect individual preferences regarding education: school environments are part of individuals' cultural background[11]. As discussed in the previous section, marginal costs of schooling are determined by a very broad set of family and cultural influences so controlling only for skills would not be enough to estimate the relative importance of short term financial constraints.

[11] Educational background variables could also be capturing the heterogeneous effect of skills in different skill groups. We added quadratic terms for skills and dummies allowing for different effects depending on the position in the skill distribution. None of these were significant and the results shown in panel D were robust to these specifications.

### 5.2 Financial constraints, skills and family background: simulation results

Figures 5 and 6 summarize the results of the simulation exercise described above. Each of the values reported in Figure 5 correspond to the values $\%\mathrm{G1}_{\mathrm{g}}$ for $g = 1,2,\dots,5$ as described in (5). Similarly, values reported in Figure 6 correspond to the values of $\%\mathrm{G2}$ for $g = 1,2,\dots,5$ as described in (6). Thus, Figure 5 reports the relative contribution of each covariate difference when explaining the gap between observed and full access. Figure 6, on the other hand, reports the relative contribution of each covariate difference when explaining the regressive nature of college access.

**Figure 5**

Percentage of the gap between observed and full (100%) higher education access closed by each covariate group

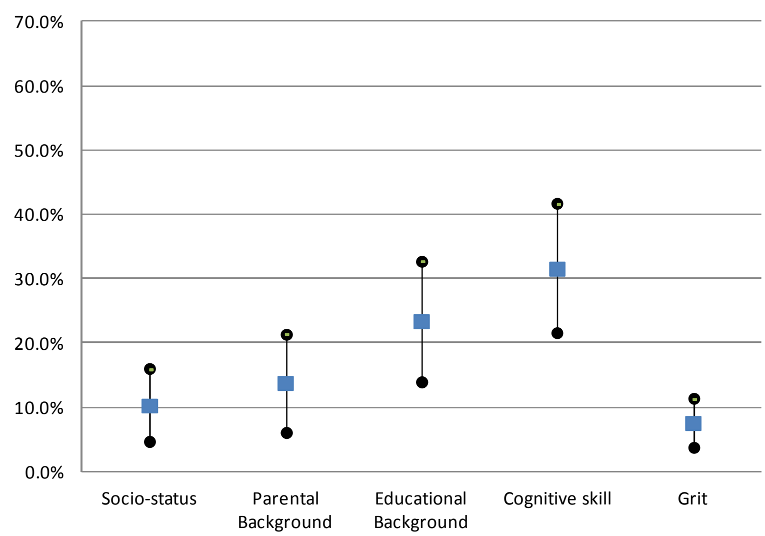
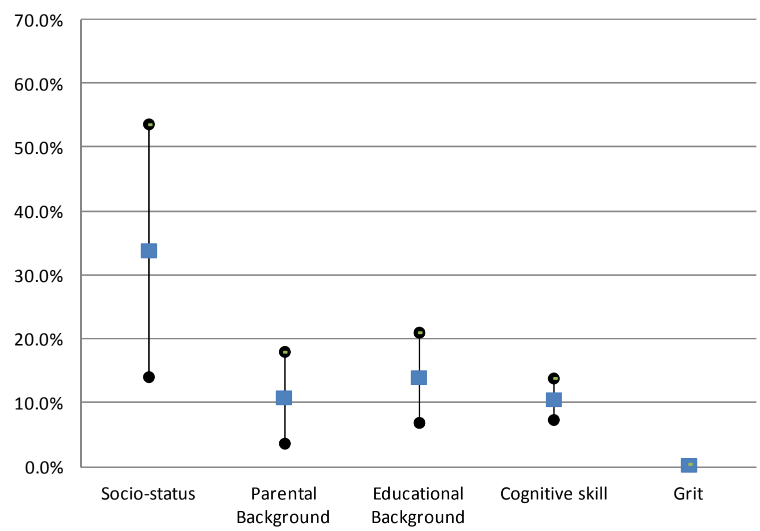


**Figure 6**
Percentage of the college access gap between low and
medium-high socioeconomic status individuals closed by each covariate group




If we seek to explain what differences between poor and non-poor individuals account for the regressive pattern of access to college education in Peru, family income explains, at most, half of the gap. As shown in Figure 6, the proportion of the college access gap between low and medium-high socioeconomic status individuals that can be related to differences in family income is around 35%, with an upper bound close to 55%. Our simulations also evidence that the rest of the gap is related to differences in parental education, educational background and cognitive skills, with similar contributions.

If we seek to explain what differences between an average individual and those who did enroll drive access to higher education in Peru, cognitive skills have a leading role and account for 30% of the gap. Differences in family income, on the other hand, account for only 10% the gap.

From the above figures is clear that the relative contribution of family income and skills is not the same when explaining higher education access (on average) or the regressive nature of college enrollment. This result should not be surprising. It is driven by the fact that differences in terms of family income between low and medium-high income individuals are stronger than between an average individual and those who did enroll in higher education. The opposite is true for differences in terms of skills: they are stronger between an average individual and those who had access to higher education, than between low and medium-high income individuals.

## 6. Concluding remarks and implications for policy

We have explored the determinants of postsecondary trajectories in urban Peru using a novel household survey on a national sample of working-age population that includes special modules of tests to measure cognitive and non-cognitive skills. Our analysis has focused on analyzing the relative importance of short term financial constraints *vis a vis* skills and other background factors affecting "tastes for education" when explaining access to higher education in Peru and the disparities between rich and poor households regarding access to college.

Our results show that family income has a role when explaining access to higher education but individual skills and educational and family background variables are also significant. Cognitive skills (numeracy and problem-solving ability, working memory, verbal fluency and receptive language) and grit (perseverance) have a significant effect on college access and, even after controlling for these skills, parental education and variables reflecting scholastic achievement and type of school (private or public) also exhibit an important effect. This conforms with the notion that family and cultural background influence tastes for education and determine the marginal cost of schooling.

If we refer to the regressive nature of college access, our simulations show that the strong correlation between college enrollment and family income in urban Peru is not only driven by credit constraints, but also by poor college readiness in terms of cognitive skills and by poor family and educational backgrounds affecting preferences for schooling. In fact, family income explains, at most, half of the college access gap between poor and non-poor households. The other half is related to differences in parental education, educational background and cognitive skills.

An important policy implication of the above is that credit and/or scholarship schemes alone will not suffice to reverse the strong regressive nature of college enrollment in

Peru. In fact, our results confirm that such policies would entail a significant equity-efficiency trade-off: transferring monetary resources to foster higher education enrollment among the poor faces the risk of focusing on a population where constraints in terms of skills and poor family and educational backgrounds are stronger. Efficiency losses could come in the form of attrition or a decline in the quality of college education. Credit and/or scholarship schemes should be accompanied by a rigorous selection process (based on skills) to ensure an efficient use of public resources. If adequately focalized, thus, they could only have limited coverage.

Early investments in the development of basic skills, on the other hand, are more difficult to deliver and its results take longer to materialize. They, however, face no equity-efficiency trade-offs (Cunha and Heckman, 2008) and, according to our results, will have a significant effect on average higher education enrollment.

In terms of future research, the rich survey data from Peru could also be used to study the role of cognitive and non-cognitive skills when explaining completion of post-secondary schooling. In addition, the effect of higher education experiences on skills (especially non-cognitive skills) could also be addressed. As already discussed, further analysis of potential attenuation and positive biases arising from measurement error and reverse causality from schooling to test scores would also be of interest. For this, longitudinal data designs are a necessary second step in data collection on skills in developing countries.

## References


Barrick, M. R., M. K. Mount, and T. A. Judge (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9(1-2)* 9-30.

Barrick, M. R. and M. K. Mount (2005). Yes, Personality Matters: Moving on to More Important Matters. *Human Performance, 18(4)* 359-372.

Bassi, M. y S. Galiani (2010). "Labor Market Insertion of Young Adults in Chile". Mimeo. Inter-American Development Bank.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. National Bureau of Economic Research, New York. Distributed by Columbia University Press.

Borghans, L., A. Duckworth, J. Heckman, and B. Weel (2008), "The Economics of Psychology and Personality Traits", IZA DP No. 3333.

Bowles, S., and H. Gintis (1976). Schooling in Capitalist America: Educational Reform the Contradictions of Economic Life. Basic Books, New York.

Cameron, S.V., Heckman, J.J. (2001). "The dynamics of educational attainment for black, Hispanic, and white males". *Journal of Political Economy* 109 (3), 455–499.

Card, D. (1994). "Earnings, Schooling and Ability Revisited". National Bureau of Economic Research Working Paper No. 4832.

Card, D. and A. Krueger (1996). "Labor Market Effects of School Quality: Theory and Evidence". National Bureau of Economic Research Working Paper No. 5450.

Carneiro, P., Heckman, J.J. (2002). "The evidence on credit constraints in post-secondary schooling". *Economic Journal* 112 (482), 705–734.

Castro, J. F. and G. Yamada (2010). "Educación superior e ingresos laborales: estimaciones paramétricas y no paramétricas de la rentabilidad por niveles y carreras en el Perú", DD/10/06, Universidad del Pacifico.

Checchi, D. (2005). The Economics of Education. Cambridge University Press, New York.

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

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

Cunha, F. and J. Heckman (2008). "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Journal of Human Resources*, 43:738-782.

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

Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R. (2007). "Grit: Perseverance and passion for long-term goals", *Personality Processes and Individual Differences*, 92 (6), p. 1087.

Eccles, J. and A. Wigfield, 2003, "Schooling's Influences on Motivation and Achievement" in Danziger and Waldfogel, eds. 2000. Securing the Future: Investing in Children from Birth to College. New York: Russell Sage Found.

Farkas, G. 2003. Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes. Annual Review of Sociology, Vol. 29, pp. 541-562

Goldberg, L. R. (1990). "An alternative "description of personality": The big-five factor structure", *Journal of Personality and Social Psychology* 59 (6): 1216–1229.
Greene, W. (2000), Econometric Analysis, Fourth Edition, Prentice-Hall.

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

Hanushek, E. and L. Woessmann (2009). "Schooling, Cognitive Skills, and the Latin American Growth Puzzle", NBER Working Paper 15066.

Heckman, J.J. (2000). "Policies to foster human capital". Research in Economics 54 (1), 3–56. With discussion.

Heckman, J.J., Rubinstein, Y. (2001). "The importance of noncognitive skills: Lessons from the GED testing program". *American Economic Review* 91 (2), 145–149.

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

Heckman, J.J., Savelyev, P., Yavitz, A. (2006). "The Perry preschool project: A reanalysis". Unpublished manuscript, University of Chicago, Department of Economics.

Heckman J. J. and F. Cunha (2010). Investing in our Young People. Mimeo.

Humlum, M., K. Kleinjans, and H. Nielsen (2007), "An Economic Analysis of Identity and Career Choice", IZA DP No. 3120.

Morón, E., J. F. Castro and C. Sanborn (2009). "Helping Reforms Deliver Inclusive Growth in Peru", in Rojas-Suarez, L. (ed.), Growing Pains in Latin America, Center for Global Development (CGD).

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313-345.

Shonkoff, J.P., Phillips, D. (2000). From Neurons to Neighborhoods: The Science of Early Child Development. National Academy Press, Washington, DC.

Spence, A.M. (1973). "Job market signaling". *Quarterly Journal of Economics* 87 (3), 355–374.

Sternberg, R.J. (1985). Beyond IQ: A Triarchic Theory of Human Intelligence. Cambridge University Press, New York.

Todd, P. and K. Wolpin (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement", *The Economic Journal*, Vol. 113: F3-F33.

**Appendix 1**

**ENCUESTA NACIONAL DE HABILIDADES (ENHAB)**
**DE PERU URBANO, 2010**

**Skills Measurement**

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

## Appendix 1 (Cont.)

## MEASURING NON-COGNITIVE TRAITS: BIG-FIVE PERSONALITY FACTORS

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

## Appendix 2

### (A) Multinomial logit estimates for the complete model ("Dit not enroll" = baseline category)

| Covariate groups | Technical | College |
|---|---|---|
| ***Socioeconomic status*** | | |
| Medium or high = 1 | 0.41** | 0.869*** |
| ***Cognitive skills*** | | |
| Aggregate measure | 0.557*** | 1.02*** |
| ***Non-cognitive skills*** | | |
| Grit | 0.05 | 0.445*** |
| Extraversion | 0.00 | -0.264* |
| Agreeableness ("easy going") | 0.12 | 0.07 |
| Agreeableness ("reliable") | -0.242** | -0.02 |
| Conscientiousness | 0.02 | -0.17 |
| Emotional stability | 0.08 | 0.03 |
| Openness | 0.14 | -0.01 |
| ***Parental background*** | | |
| Father educational attainment (secondary = 1) | 0.604** | 0.558** |
| Father educational attainment (higher = 1) | 0.676** | 1.326*** |
| Mother educational attainment (secondary = 1) | -0.16 | -0.28 |
| Mother educational attainment (higher = 1) | 0.06 | 0.10 |
| Importance given by parents to education (high = 1) | 0.03 | -0.19 |
| Importance given by mother to education (high = 1) | -0.52 | 0.682** |
| ***Educational background*** | | |
| Preschool (public = 1) | -0.511** | -0.373* |
| Preschool (private = 1) | -0.18 | -0.76 |
| Public school = 1 | 0.07 | -1.096*** |
| Had to repeat a year or more in school = 1 | -0.27 | -0.76*** |
| Perception regarding performance (top student = 1) | 0.45 | 0.984*** |
| Perception regarding effort (large = 1) | 0.497*** | 0.525** |
| | | |
| Constant | -1.017* | -0.866° |

Number of obs. = 1,674

Pseudo R2 = 0.2164

Significant at: 1% (***), 5% (**), 10% (*), 15% (°)

All models control for age, sex, first language, birth order, number of siblings, and born in Lima (capital city).

## (B) Bivariate probit estimates used to test for selection within the sample of high school graduates

| **Covariate groups** | **Pr(complete high school)** | **Pr(Enrolled in higher education \| complete high school)** |
|---|---|---|
| ***Socioeconomic status*** | | |
| Medium or high = 1 | 0.50*** | 0.27** |
| ***Cognitive skills*** | | |
| Aggregate measure | 0.56*** | 0.41*** |
| ***Non-cognitive skills*** | | |
| Grit | -0.05 | 0.11* |
| Extraversion | 0.08 | -0.07 |
| Agreeableness ("easy going") | 0.02 | 0.03 |
| Agreeableness ("reliable") | -0.11 | -0.05 |
| Conscientiousness | 0.04 | -0.04 |
| Emotional stability | 0.01 | 0.01 |
| Openness | 0.15* | 0.04 |
| ***Parental background*** | | |
| Father educational attainment (secondary = 1) | 0.34** | 0.32** |
| Father educational attainment (higher = 1) | 0.69** | 0.62*** |
| Mother educational attainment (secondary = 1) | 0.00 | -0.11 |
| Mother educational attainment (higher = 1) | 0.06 | 0.03 |
| Importance given by parents to education (high = 1) | 0.27** | -0.12 |
| Importance given by mother to education (high = 1) | 0.08 | 0.09 |
| ***Educational background*** | | |
| Preschool (public = 1) | -0.30* | -0.29** |
| Preschool (private = 1) | 0.12 | -0.29 |
| Public school = 1 | 0.21 | -0.31 |
| Had to repeat a year or more in school = 1 | -0.32*** | -0.2 |
| Perception regarding performance (top student = 1) | 0.44 | 0.39** |
| Perception regarding effort (large = 1) | 0.09 | 0.27** |
| | | |
| Constant | -0.17 | -0.01 |
| | | |
| **Wald test of indep. eqns.** | | |
| Prob > chi2 = 0.3111 | | |

Number of obs. = 1,876

Significant at: 1% (***), 5% (**), 10% (*), 15% (°)

All models control for age, sex, first language, birth order, number of siblings, and born in Lima (capital city).

**Appendix 3: Potential attenuation and positive biases when estimating the effect of skills on higher education enrollment**

For simplicity, let us assume a linear probability model for higher education enrollment.

$$E_i = \alpha_0 + \alpha_1 A_i + x_i'\delta + \varepsilon_i \qquad \text{(i)}$$

Where $A_i$ is individual ability, $x_i'$ contains other controls affecting enrollment, and $\varepsilon_i$ is a random shock affecting enrollment which is uncorrelated with ability and variables in $x_i'$. Trying to measure the relative importance of skills for higher education enrollment implies searching for a consistent estimate of $\alpha_1$.

As discussed in the main text, test scores are not ability but instead reflect *measured ability* ( $MA_i$). Measured ability is, of course, a function of ability but can also be affected by higher education and measurement error.

$$MA_i = A_i + \gamma E_i + \mu_i \qquad \text{(ii)}$$

The above implies the following relation between enrollment and measured ability.

$$E_i = \alpha_0 + \alpha_1( MA_i - \gamma E_i - \mu_i) + x_i'\delta + \varepsilon_i$$

$$E_i(1 + \alpha_1\gamma) = \alpha_0 + \alpha_1 MA_i + x_i'\delta + \varepsilon_i - \alpha_1\mu_i \qquad \text{(iii)}$$

Thus, in and empirical specification where enrollment is regressed on test scores we have:

$$E_i = \beta_0 + \beta_1 MA_i + x_i'\theta + \omega_i \qquad \text{(iv)}$$

where $\beta_1 = \alpha_1/(1 + \alpha_1\gamma)$ and $\omega_i = (\varepsilon_i - \alpha_1\mu_i)/(1 + \alpha_1\gamma)$.

To keep the algebra simple, let us abstract from the presence of covariates contained in $x_i'$ (or assume they are orthogonal to measured ability). If we denote as $\widetilde{M}A_i$ and $\widetilde{\omega}_i$ the corresponding variables deviated from their sample means, the OLS estimate of $\beta_1$ can be expressed as:

$$\hat{\beta}_{1,OLS} = \beta_1 + \frac{\sum_i( \widetilde{M}A_i)(\widetilde{\omega}_i)}{\sum_i( \widetilde{M}A_i)^2} \qquad \text{(v)}$$

And its probability limit can be solved as follows.

$$plim\ \hat{\beta}_{1,OLS} = \beta_1 + \frac{Cov(MA_i, \omega_i)}{Var(MA_i)} \qquad \text{(vi)}$$

$$plim\ \hat{\beta}_{1,OLS} = \beta_1 + \frac{E\{[A_i + \gamma E_i + \mu_i][(\varepsilon_i - \alpha_1\mu_i)/(1+\alpha_1\gamma)]\}}{Var(MA_i)}$$

$$= \beta_1 + \frac{1}{\sigma_{MA}^2(1+\alpha_1\gamma)}\left[\gamma\sigma_\varepsilon^2 - \alpha_1\sigma_\mu^2\right]$$

$$= \beta_1\left[1 - \frac{\sigma_\mu^2}{\sigma_{MA}^2}\right] + \frac{\gamma\sigma_\varepsilon^2}{\sigma_{MA}^2(1+\alpha_1\gamma)}$$

$$= \frac{\alpha_1}{(1+\alpha_1\gamma)}\left[1 - \frac{\sigma_\mu^2}{\sigma_{MA}^2}\right] + \frac{\gamma\sigma_\varepsilon^2}{\sigma_{MA}^2(1+\alpha_1\gamma)} \tag{vii}$$

Clearly, the possibility of obtaining a consistent estimate of $\alpha_1$ from $\beta_{1,OLS}$ depends on higher education enrollment having no effect on measured ability $(\gamma = 0)$ **and** a small noise to signal ratio $\left(\sigma_\mu^2/\sigma_{MA}^2 = 0\right)$.

We can implement an IV strategy to try to purge the effect of higher education enrollment on measured ability and use the "residualized" version of measured ability $(RMA_i)$.

$$\begin{aligned} RMA_i &= MA_i - \hat{\gamma}_{IV}E_i \\ &= A_i + E_i(\gamma - \hat{\gamma}_{IV}) + \mu_i \end{aligned} \tag{viii}$$

Our empirical specification has now the form:

$$E_i = \rho_0 + \rho_1\ RMA_i + x_i'\varphi + v_i \tag{ix}$$

where $\rho_1 = \alpha_1/\left(1 + \alpha_1(\gamma - \hat{\gamma}_{IV})\right)$ and $v_i = (\varepsilon_i - \alpha_1\mu_i)/\left(1 + \alpha_1(\gamma - \hat{\gamma}_{IV})\right)$. If denote $plim\ (\gamma - \hat{\gamma}_{IV}) = \vartheta$ and follow the same procedure as above, the probability limit of $\hat{\rho}_{1,OLS}$ can be solved as:

$$plim\ \hat{\rho}_{1,OLS} = \frac{\alpha_1}{1+\alpha_1\vartheta}\left[1 - \frac{\sigma_\mu^2}{\sigma_{RMA}^2}\right] + \frac{\vartheta\sigma_\varepsilon^2}{\sigma_{RMA}^2(1+\alpha_1\vartheta)} \tag{x}$$

A successful IV strategy implies $\vartheta = 0$ and, as discussed in the main text, we would still have an attenuation bias preventing a consistent estimation of $\alpha_1$. If we are unable to isolate variation in higher education enrollment uncorrelated with skills and overestimate the effect of college attendance on measured ability, we will have $\vartheta < 0$ and our estimate of $\alpha_1$ could even be negative.