Transition from School to Work: The Role of Cognitive and Noncognitive Abilities

Duc Thanh Nguyen*

June 10, 2024

Abstract

This paper examines the effects of cognitive and noncognitive skills on college completion decisions and subsequent earnings. I explicitly embed a model of endogenous education decisions and subsequent earnings into a latent factor model. This approach allows for the identification of latent competencies to capture multiple skill dimensions more accurately and correct for measurement errors in observed measures of skills. It also allows for the isolation of the effects of these skills on earnings into components explained by schooling and productivity. Furthermore, this approach solves the endogeneity and reverse causality problems of skills, schooling and earnings by excluding education variables from earnings equations, introducing latent skills and using panel data with skills and outcomes observed at different times. The findings indicate that both cognitive and noncognitive skills in adolescence are associated with college completion and better earnings in early adulthood. Both types of skills are important in directly determining earnings and indirectly determining earnings through their influence on schooling.

Keywords: Cognitive skills, noncognitive skills, school-to-work transition, schooling, earnings, college completion, dynamic factor analysis, latent factor model.

*Department of Economics, Concordia University, e-mail: *ducthanhht@gmail.com*

1 Introduction

Schooling has been widely used as a proxy for understanding the impact of skills on labor market outcomes. However, evidence has shown that educational attainment is insufficient to ensure labor market success since it does not necessarily guarantee the required knowledge and skills. Schooling generally raises wages only if it generates skills that create labor productivity and have returns in the labor market (Hanushek, 2002). It is especially true when skills vary widely for children with similar schooling levels (Hanushek and Woessmann, 2008; Singh, 2019). Identifying and understanding the effects of cognitive and noncognitive skills on college education and labor outcomes is vital to better inform policy designs for which skills are rewarded by the labor market and should be improved. However, one of the key challenges in assessing the impact of skills is that it is difficult to reliably capture multiple dimensions of skills with several imperfect candidate measures in surveys. This study aims to model latent skills as a source of unobserved heterogeneity to capture the dimensions of skills more accurately and then assess the effects of these skills on endogenous educational choices and subsequent earnings.

Recently, researchers have paid increasing attention to more accurate skill measures and their impacts on labor market outcomes. These skills can be classified into two categories, cognitive skills and noncognitive skills. Research has shown that cognitive and noncognitive skills play a critical role in educational attainment, labor market success as well as the development of successful lives (Almlund et al., 2011; Cunha and Heckman, 2009; Heckman et al., 2006; Urzúa, 2008; Cawley et al., 2001). Most researchers have focused on the role of cognition in explaining schooling and economic outcomes. Recently, the literature has focused on the impact of noncognitive skills on education and labor productivity, but evidence on the impacts of noncognitive skills on schooling and labor market outcomes is still much limited. While there is consensus on the important role of both types of skills, the relative importance of these skills is still debated. Understanding what skills are required and how specific skills are rewarded by the market is key for designing policies to improve labor productivity and reduce inequality in the long run.

In assessing the determinants and roles of cognitive and noncognitive skills, most research directly uses measured test scores as a proxy for cognitive and noncognitive skills. However, observable tests such as math or IQ tests are not perfect measures of abilities since they suffer from measurement errors. Furthermore, dependence across skills, schooling and labor market outcomes gives rise to the problems of reverse causality and endogeneity.

Vietnam is an interesting case study, not only because panel data is available but also from a policy point of view. Vietnam has achieved significant success in the education sector and its results in students' acquisition of cognitive skills are especially impressive. Vietnam was ranked 12th out of 76 participating countries in the Program for International Student Assessment (PISA) in 2012, above the OECD average as well as above the United States, Australia, and the United Kingdom. However, labor productivity in Vietnam is among the lowest in the Asia-Pacific region (GSO, 2016). One possible explanation for the paradox of very high levels of cognitive skills and educational achievement but very low productivity is low noncognitive skills and the lack of the right skills to match the labor market. Despite impressive achievements in cognitive skills, most of employers faced a shortage of workers with the required skills and unavailability of skilled applicants (Bodewig et al., 2014). Failure in equipping students with adequate noncognitive skills is one prominent weakness of the Vietnamese formal education and training system. The Vietnamese labor force's scores for noncognitive skills are low compared to developing countries (Roseth et al., 2016) and the formal education sector in Vietnam has provided the workforce with low 'soft' skills (Bodewig et al., 2014). In the context of the increased demand for high-skilled workers and the decreased demand for low-skilled workers as a result of rapid technological development and economic integration, strong economic growth will require Vietnam to hinge on the higher productivity of labor, which is substantially influenced by skill shortages and mismatches. While the labour force is equipped with high cognitive but low noncognitive skills, there is a shortage of workers with high noncognitive skills. The shortage of the right skills including non-cognitive skills and mismatches of skills can cause lower productivity. The evidence for the relative importance of skills in determining labor outcomes is thus crucial for suggesting which type of skills should be enhanced more intensively to improve labor outcomes, inclusive of earning capacity or labor productivity.

This study contributes to the limited evidence about the effects of different skills, especially noncognitive skills, on labor market outcomes in developing countries that identify unobserved heterogeneity, correct problems of measurement errors, reverse causality and endogeneity. First, I use a structural latent factor approach to identify true or latent skills as a source of unobserved heterogeneity and their distributions instead of employing (noisy) proxy observed skill measures to capture more accurately multiple cognitive and noncognitive skill dimensions and correct any measurement errors in observed skills. Second, I explicitly embed a model of endogenous education decisions and subsequent earnings into a latent factor model. This strategy solves the endogeneity and reverse causality problems of skills, schooling and earnings by excluding education variables from earnings equations, introducing latent skills and using longitudinal data with skills and outcomes observed at different times. Finally, the model allows schooling decisions to be endogenous to examine the effects of skills on college completion decisions and isolate the effects of these skills on earnings into components explained by schooling and productivity. This approach is crucial to understand the effects of skills and college education on labor outcomes.

2 Literature

Panel data are needed to assess the impacts of skills (cognitive and noncognitive skills) on schooling and labor market outcomes. Studies using cross-section data are subject to a risk of reverse causality between skills, schooling and labor market outcomes because they are observed simultaneously, and schooling and work experience may greatly influence skills. However, longitudinal data that measures both cognitive and noncognitive skills during childhood and follows those children into adulthood is rare. Most studies on skills' impacts on schooling and labor market outcomes are on developed countries, in particular, the United States; there is little such evidence in developing countries.

Studies have shown that both cognitive skills and noncognitive skills affect schooling decisions (Almlund et al., 2011) and labor market outcomes (Hanushek and Woessmann, 2008; Almlund et al., 2011; Hanushek, 2009, among others) and cognitive skills have relatively more significant effects than noncognitive skills.

Cunha et al. (2010) show that cognitive and noncognitive skills account for 16 and 12 percent of the variance in educational attainment respectively in the US. Mathematics, reading, and attention skills strongly influence educational success, while noncognitive skills have a limited impact on educational outcomes in the United Kingdom, the United States, and Canada (Duncan et al., 2007).

A large body of evidence has shown that higher cognitive skills measured by test scores such as mathematics, reading, and vocabulary were associated with higher incomes (Murnane et al., 2000; Cawley et al., 2001; Green and Riddell, 2003; Hanushek and Woessmann, 2008; Heckman et al., 2006; Hanushek and Zhang, 2009; Hanushek et al., 2015). For example, studies in the US find that a one standard deviation increase in the 12th-grade math test score increases annual earnings by 10−15 percent (Murnane et al., 2000; Lazear, 2003). A one standard deviation increase in literacy scores increases earnings by 9.3 percent in a 13-country sample (Hanushek and Zhang, 2009).

Evidence on the relationship between noncognitive abilities, schooling and economic outcomes is much scarcer. A newly growing literature shows that noncognitive competencies have equally important effects as cognitive abilities on schooling and labor market outcomes (Heckman et al., 2006; Cunha and Heckman, 2008; Lindqvist and Vestman, 2011; Almlund et al., 2011). A review of 13 studies by Lindqvist and Vestman (2011) indicates that a one standard deviation increase in noncognitive abilities would increase wages by 4 to 8%. Heckman et al. (2006) show that noncognitive skills are as equally important as cognitive skills in explaining labor wages in the US and increasing noncognitive skills by one standard deviation would increase wages by 11.2%. Using the same data, Heckman et al. (2011) show that skills strongly impact educational attainment and influence earnings through their effects on education, but given years of schooling, noncognitive skills have little direct effects on wages.

Numerous studies directly use test scores as a proxy for cognitive and noncognitive skills (Long et al., 2015; Nordman et al., 2015; Sahn and Villa, 2015; Krishnan and Krutikova, 2013; Díaz et al., 2012 among others). However, ability is multidimensional; it depends not only on skills, but also on other factors and the dimensions of the skill set measured in the survey. Test scores are difficult to measure precisely and are noisy proxies for underlying cognitive and noncognitive abilities; using test scores as a proxy for skills suffers from measurement errors. Furthermore, these studies suffer from the problem of endogeneity and reverse causality. Endogeneity may arise when education is included in earnings equations and reverse causality arises when using cross-section data with skills and outcomes observed simultaneously. Most of the existing evidence address the association rather than causality in making inference on the effects of skills on labor market outcomes. In particular, in this approach, wages depend on schooling choices, observed cognitive and noncognitive test scores and other controls. However, schooling depends on these scores, so schooling is endogenous in wage estimations. Moreover, higher wages could affect schooling choices, and schooling at the time of tests also affects cognitive and noncognitive test scores. This causes a problem of reverse causality between schooling and wages and between schooling and observed test scores; and test scores are endogenous. Omitting schooling variables from wage equations can solve the endogeneity problem of schooling. However, this only allows us to estimate the net effects of skills, but does not fully capture the indirect effects of skills on wages via schooling and test scores are still endogenous.

Heckman et al. (2006), Cunha et al. (2010), and Heckman et al. (2011) develop and use structural measurement frameworks to address measurement errors in measuring skills and the endogeneity of observed skills. Murnane et al. (2001) and Drago (2011) use skills measured before individuals enter the job market to address reverse causality, but they have not addressed measurement errors in test scores and have not taken into account the endogeneity of schooling or investments in estimating the effects of skills.

Evidence on the impacts of cognitive and noncognitive skills for developing countries is rare because surveys measuring both cognitive and noncognitive skills during childhood and following those children into adulthood were unavailable and data sets are mainly cross-sectional data on adults' cognitive skills and noncognitive skills which are primarily related to the job. New data from developing countries allows the exploration of whether skills are as important in labor markets as in developed countries and allow for a causal identification strategy. Studies in developing countries show that both types of skills predict schooling and wages. While most of these studies (Cunningham et al., 2016; Nordman et al., 2015; Sahn and Villa, 2015; Acosta et al., 2015; Krishnan and Krutikova, 2013; Díaz et al., 2012) give the first sets of evidence on how these skills influence schooling decisions and labor market outcomes, they suffer from either the problem of reverse causality or endogeneity bias. Cunningham et al. (2016) and Acosta et al. (2015) deal with measurement errors in measuring skills. However, these two papers use cross-sectional data to examine the effects of cognitive and noncognitive skills of adults on contemporaneous labor market outcomes and they examine the direct impacts of skills. Endogeneity and reverse causality issues between skills and labor market outcomes may still remain in their cross-sectional data studies. Evidence in both developed and developing countries shows that skills affect labor outcomes not only directly but also indirectly through their effects on schooling (Heckman et al., 2006, Heckman et al., 2011; Glewwe et al., 2017). Furthermore, they use skills measured when subjects were adults instead of adolescent skills as in this study.

3 Data and Definition

The Vietnam Young Lives survey follows 2000 children in the Younger Cohort and 1,000 children in the Older Cohort. There are five rounds of the survey. The first round was conducted in 2002, at the age of 1 for the Younger Cohort and 8 for the Older Cohort, followed every 4 years until age 15 and 22 for the Younger and Older Cohorts respectively.

The Young Lives survey provides a rich set of data on diversified aspects of children, their families and communities. The survey includes the child, household and commune questionnaires and collects comprehensive information on individual, family, caregiver and parent characteristics and resources, their preferences and feelings as well as schools and communities. This study uses the Young Lives survey data for the Older Cohort that measures both cognitive and noncognitive skills during childhood and the survey follows those children into adulthood with high-quality information on their schooling and labor market outcomes. The data set also provides rich information on children's surrounding environment. This rich available information over time enables us to study skill formation and the impacts of skills on adult outcomes and it allows us to solve the problems of reverse causality and endogenously between skills, schooling and labor market outcomes. In this study, I use skills measured at the age of 15 and assess their impact on the decision to complete college and market outcomes at the age of 22.

Although the Young Lives survey sampling is not designed to be nationally representative of the population, it covers the diversity of children in the country in terms of a wide variety of attributes and experiences. The diversity of children allows us to analyze causal relations and the changing dynamics of childhood welfare over time. As a longitudinal survey, Young Lives is intended to show changes for individuals over time and the impact of earlier circumstances on children's later outcomes. This survey uses a sentinel-site sampling design comprising 20 purposely selected sites chosen to represent diversity, but with a pro-poor bias (Nguyen, 2008). At the site level, children were selected randomly in 2001 such that the data are representative of the birth cohort at each site. My analysis will be conducted for those who completed all cognitive and noncognitive tests and those with complete education and income data at age 22.

As with any longitudinal survey, sample attrition is always an issue. The Young Lives survey is concerned to minimize attrition. The attrition rate for the Young Lives survey for the Older Cohort in Vietnam is 9% since the start of the survey and it is relatively low compared to the other study countries and other longitudinal surveys.[1] Given that I examine cognitive and non-cognitive skills and early earnings, the panel sample is restricted to include those individuals who have complete skills, schooling and earnings data and the final panel consists of 757 observations.[2]

3.1 Cognitive Skills

Cognitive skills, also called cognition, cognitive abilities or intelligence, can be simply defined as knowledge and one's ability to acquire new knowledge (Glewwe et al., 2017). VandenBos (2007), in the American Psychological Association Dictionary of Psychology, defines cognitive skills as "all forms of knowing and awareness, such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving". Cognitive skills are normally measured by cognitive test scores.

[1] The cumulative attrition rates are 1%, 2.4%, 11.3 % and 9% in Round 2, 3, 4 and 5 respectively.

[2] Similar studies using data from the US, Canada and other countries drop many observations because of the sample restrictions. For example, Prada and Urzúa (2017) end up with 1,022 out of 12,686 observations from an original sample; Heckman et al. (2006) end up with 4680 out of 12,686 observations and Kottelenberg and Lehrer (2019) end up with 1,607 out of 29,687 individuals.

The cognitive development in the Young Lives survey is measured by the test scores in mathematics (Math test), reading comprehension (Cloze test), Peabody Picture Vocabulary Test (PPVT) and Language test (Vietnamese). The Math test and Peabody Picture Vocabulary Test were administered to the Older cohort from Round 1 to Round 3. The Cloze test was added in Round 3; and in Round 4, the PPVT test was replaced by the Language test (in Vietnamese).

Math test: The Math test was administered in Rounds 2 and 3. It includes 29 items on addition, subtraction, multiplication, division, problem-solving, measurement, data interpretation, and basic geometry.

PPVT: The PPVT is a widely-used test of receptive vocabulary. It uses a stimulus word and accompanying pictures to test receptive vocabulary. It has been extensively used to demonstrate the correlation between PPVT scores and cognitive and intellectual ability (Walker et al., 2005). The 204−item PPVT−III was used in Vietnam. Young Lives researchers in each country followed a standard process for adaptation and standardization of the PPVT.

Cloze: The Cloze test was developed to measure verbal skills and reading comprehension. The test includes 24 items that increase in difficulty. Each item consists of a sentence or short paragraph that lack one or more words; children were asked to identify a word that completed the meaning of the sentence or paragraph. A thorough analysis of psychometric characteristics was examined to establish the reliability and validity of all these tests (Crookston et al., 2014).

3.2 Noncognitive Skills

Noncognitive skills, also called soft skills, social-emotional skills, noncognitive competencies, noncognitive abilities, or personality traits, can be defined as patterns of feelings, thoughts and behaviors (Borghans et al., 2008; Thiel and Thomsen, 2013).

This study uses three composite indicators designed to access dimensions of self-esteem, self-efficacy and self-respect and inclusion to measure noncognitive skills.

Self-esteem: Self-esteem measures aspects related to pride and it builds on the Rosenberg scale (Rosenberg, 1965). It is related to a person's overall evaluation of their worth. The statements used to measure self-esteem are adapted from the Rosenberg Self-Esteem Scale and focus on different dimensions of the child, such as housing, clothing, work and school.

Self-efficacy: the self-efficacy scale measures aspects related to agency and builds on the Rotter scale (Rotter, 1966). It is related to a person's sense of agency or mastery over his life. The statements used to measure self-efficacy focus on different domains of the child, such as school, work and time use.

Self-respect and inclusion: focus on the social component of self-esteem (Dercon and Krishnan, 2009). The statements used to measure self-respect revolve around the concepts of pride and sense of inclusion.

The measures of self-esteem, self-efficacy and self-respect and inclusion are set on a five-point Likert scale ranging from "strongly disagree" to "strongly agree". Children were read statements and asked whether they strongly disagreed, disagreed, more or less, agreed or strongly agreed with the statements. Negative statements are recoded to reflect positive statements. The self-esteem index includes six items/statements, the self-efficacy index consists of five items and the self-respect and inclusion contain nine items. Each item is standardized with mean 0 and variance 1 and the three composite indices of the noncognitive skills - self-esteem, self-efficacy and self-respect and inclusion - are the average of standardized items used to construct each index. The aim is to place all measurements on the same scale and approximate a measure associated with values of the psychosocial competencies (Dercon and Krishnan, 2009).

The statements used to construct self-esteem, self-efficacy and self-respect and inclusion indices were drawn from the educational psychology literature, they were adapted and extensively tested during piloting to apply for children across different cultures (Dercon and Sánchez, 2013). Self-esteem and self-efficacy are the most popular noncognitive skill measures used in empirical studies (Glewwe et al., 2017). Self-respect and inclusion are related to the self-esteem measure but focus on social and psychosocial aspects of inclusion. Of personality traits, these indices have been found to strongly predict educational achievements and adult social and economic outcomes (Almlund et al., 2011) and they have been used in numerous studies in Vietnam (Dercon and Krishnan, 2009; Dercon and Sánchez, 2013; Sánchez, 2013; Sánchez and Singh, 2018; Singh, 2019).

Table 1: Scores Used to Measure Cognitive and Noncognitive Skills

Skills	**No. of items**
Cognitive Skills	
1. Peabody Picture Vocabulary Test (PPVT)	204
2. Mathematics Test (Math Test)	29
3. Cloze Test	24
Noncognitive skills:*	
1. Self-esteem Scale	6
2. Self-efficacy Scale	5
3. Self-respect and Inclusion	9

* The items/statements used to construct composite noncognitive measures (Self-esteem scale, Self-efficacy Scale, Self-respect and Inclusion) are detailed in Appendix A.

An exploratory factor analysis of individual skills is conducted to find whether there are factors that represent cognitive skills and noncognitive skills. The factor analysis results for cognitive and noncognitive skills are provided in Appendix B. The outputs from the factor analysis based on both Kaiser's eigenvalue rule shown in Appendix Tables B.1 and B.3 and scree tests displayed in Figures B.1 and B.2 indicate that there is one factor that should be extracted from the measures of cognitive skills and one factor should be extracted from the measures of noncognitive skills.

3.3 Data on Education and Earnings

Young Lives in Vietnam collects detailed data on each child's educational outcomes, including whether the child attended kindergarten, the age when a child started primary school, the highest grade completed, the highest certificate/diploma obtained, current enrolment and detailed educational history. Table 2 presents descriptive statistics of the data. The panel sample includes 757 young people aged from 21 to 23 and is balanced between girls and boys, with 51.7% and 48.3% of females and males respectively. They have 0.613 siblings on average. People living in urban areas account for 16.9% of the sample; this reflects the pro-poor sampling approach designed by Young Lives in Vietnam (Nguyen, 2008). Girls score higher than boys in terms of both cognitive and noncognitive abilities. The average hourly earnings are 17.184 Vietnamese Dong, of which boys earn more than girls.[3] Of the total sample, 25.4% completed and obtained a college or university degree. They have 1.833 years of work experience.

4 Model

This study is built on the general framework developed by the Roy model (Roy, 1951). It models self-selection into college and potential earnings. Individuals make choices so as to maximize the potential labor outcomes based on their comparative advantages of latent talents that affects their college choices, but may not be directly applied to their job. Individuals choose a college degree based on their expected income and their own abilities.

This model follows Heckman et al. (2006), Carneiro et al. (2003), Cunha et al. (2010). The model deals with main problems in estimating the effects of skills on education and income: test scores are just proxies for true abilities and the endogeneity and reverse causality of schooling, skills and income exist in earnings equations. I estimate the model in one step. The observed measures of skills when the children were in Round 3, at the age of 15, are used to estimate the unobserved abilities - the two latent skills by a measurement system. I use the factor approach to identify these factors and their distribution rather than directly use noisy proxy variables or test scores as measures of abilities, as most of the literature does. The underlying cognitive and noncognitive skills are latent rather than observable; they are unobserved to the econometricians and are, in turn, relevant determinants of outcomes, choices and scores. Since the underlying cognitive and noncognitive factors are unobserved, I integrate over the distributions of the two latent factors and examine the effects of skills on the decision whether to complete a college degree or not that the child makes after age 15 and separate the effects of latent skills on labor market outcomes at age 22 into components explained by schooling and skills.

[3] The official exchange rate in 2016 is 21,935 Vietnamese Dong per U.S. dollar. Source: https://data.worldbank.org/indicator/PA.NUS.FCRF?locations=VN. Accessed 25 December 2022.

Table 2: Descriptive Statistics

	Full	Female	Male
PPVT	-0.091	-0.018	-0.170
	(0.947)	(0.909)	(0.982)
Math	-0.142	0.038	-0.333
	(0.948)	(0.921)	(0.940)
Cloze	-0.083	0.068	-0.247
	(1.014)	(0.927)	(1.078)
Self-Esteem	-0.023	-0.004	-0.043
	(0.639)	(0.677)	(0.597)
Self-Efficacy	-0.030	-0.001	-0.060
	(0.518)	(0.540)	(0.493)
Self-respect and inclusion	-0.021	0.021	-0.066
	(0.564)	(0.551)	(0.575)
Female	0.517	1.000	0.000
	(0.500)	(0.000)	(0.000)
Age (in years)	22.279	22.281	22.277
	(0.336)	(0.344)	(0.328)
Urban	0.169	0.174	0.164
	(0.375)	(0.380)	(0.371)
Number of siblings aged 0-18	0.613	0.708	0.511
	(0.801)	(0.824)	(0.765)
Child's educational aspiration	0.604	0.701	0.500
	(0.489)	(0.459)	(0.501)
Wealth index	0.609	0.624	0.593
	(0.177)	(0.172)	(0.181)
Parental educational level	2.392	2.494	2.284
	(1.069)	(1.079)	(1.050)
College completion	0.254	0.327	0.175
	(0.435)	(0.470)	(0.380)
Monthly earnings (1,000 VND)	3518.837	3271.102	3783.494
	(3242.325)	(3451.621)	(2984.715)
Hourly earnings (1,000 VND)	17.184	16.198	18.236
	(15.802)	(16.528)	(14.939)
Work experience (in Years)	1.833	1.724	1.949
	(1.871)	(1.819)	(1.921)
Observations	757	391	366

Note: Standard deviations in parentheses.

4.1 Measurement System

The main challenge in estimating the parameters of this model is that ability is not directly measured. It is challenging to measure ability precisely because of its multidimensional nature. Observed measures of skills or test scores should be considered only as noisy and imperfect proxies for ability, they are based on a noisy signal of one's underlying ability, and thus they suffer from measurement errors. A factor model approach allows for extracting these unobserved skills from a large set of observed data.

Cognitive skills are governed and identified by the latent factor associated with three test scores: Mathematics (T_{math}), Cloze (T_{cloze}), and PPVT (T_{ppvt}) and noncognitive skills are governed and identified by the latent factor associated with self-esteem (T_{ses}), self-efficacy (T_{sef}), and respect and inclusion (T_{ser}), in the following form:

$$T_{ij} = \alpha_j + \beta_j \theta_i^C + u_{ij} \tag{1}$$

for j = {1, 2, 3}= {math, cloze, ppvt}.

$$T_{ik} = \alpha_k + \beta_k \theta_i^{NC} + u_{ik} \tag{2}$$

for k = {1, 2, 3} = {ses, sef, ser}.

Where T_{ij} and T_{ik} are 3x1 vectors of the test scores or observed measures j and k of individual i found in the data associated with latent cognitive and noncognitive skills, θ_i^C and θ_i^{NC}, respectively. α_j and α_k are the constants. β_j and β_k are vectors of the factor loadings of the latent skills. u_{ij} and u_{ik} are error terms, which are independent of the associated factors $u_{ij} \perp \theta_i^C$, $u_{ik} \perp \theta_i^{NC}$ and they are mutually independent with an associated distribution $f_{u_h}(.)$ for h = {j, k} = {math, cloze, ppvt, ses, sef, ser}. This independence means that all the correlation in observed measures is captured by latent unobserved factors.

Specifically, the measurement system takes the following form:

$$
\begin{aligned}
T_{i,ppvt} &= \alpha_{ppvt} + \beta_{ppvt}\theta_i^C + u_{i,ppvt} \\
T_{i,math} &= \alpha_{math} + \beta_{math}\theta_i^C + u_{i,math} \\
T_{i,cloze} &= \alpha_{cloze} + \beta_{cloze}\theta_i^C + u_{i,cloze} \\
T_{i,ses} &= \alpha_{ses} + \beta_{ses}\theta_i^{NC} + u_{i,ses} \\
T_{i,sef} &= \alpha_{sef} + \beta_{sef}\theta_i^{NC} + u_{i,sef} \\
T_{i,ser} &= \alpha_{ser} + \beta_{ser}\theta_i^{NC} + u_{i,ser}
\end{aligned} \tag{3}
$$

This structure assumes that the two factors are identified by two different sets of scores. Specifically, only the latent cognitive factor is allowed to affect the individual cognitive skill scores and the latent noncognitive factor is allowed to affect the individual noncognitive skill scores or an increase in the latent cognitive factor would increase the mathematics (T_{math}), Cloze (T_{cloze}), and PPVT (T_{ppvt}) scores and any increase in the latent noncognitive factor would increase self-esteem (T_{ses}), self-efficacy (T_{sef}), and respect and inclusion (T_{ser}). That is, each measure is allocated to a dedicated factor. An alternative setting of the factors where the cognitive scores depend on both the cognitive and noncognitive factors is also considered and the results are quite similar (see Appendix E).

The distributions of the error terms u_{ih}, $f_{u_h}(.)$, are assumed to follow normal distributions with mean zero and variance $\sigma^2_{u_h}$ and let $\theta_i = \{\theta_i^C,\ \theta_i^{NC}\}$, then

$$
f(T_{ih}|\theta_i) = \frac{1}{\sqrt{2\sigma^2_{u_h}\pi}} exp\left(-\frac{(T_{ih} - \alpha_h - \beta_h\theta_i)^2}{2\sigma^2_{u_h}}\right) \tag{4}
$$

The probability of observing measures conditional on θ_i is therefore:

$$
f(T_i|\theta_i) = \prod_{h=1}^{h} f(T_{ih}|\theta_i) \tag{5}
$$

Identification of the factors requires a number of available test scores or skill indexes such that $L >= 2k + 1$, where L is the number of scores and k is the number of factors (Cunha et al., 2010; Carneiro et al., 2003). This condition of identification in this case is satisfied since there are three test scores for the cognitive factor and three indexes for the noncognitive factor. Identification also requires normalizations, I normalize one of the loadings for each factor to one and the remaining coefficients are explained in proportion to the normalized coefficients. Specifically, I set $\beta_{ppvt} = 1$ and $\beta_{ses} = 1$, thus the cognitive skill, θ^C, takes the metrics of PPVT; the noncognitive skill, θ^{NC}, takes the metrics of self-esteem. The locations of the factors are identified by setting one of the constants for each factor to zero. I set $\alpha_{ppvt} = 0$ and $\alpha_{ses} = 0$. By making these normalizations and following the identification strategy of Cunha et al. (2010), the distribution of θ for each latent skill, $F(\theta^C)$ and $F(\theta^{NC})$, and the parameters of interest are identified. I approximate the distributions of the factors by a mixture of normals, as detailed in subsection 1.4.4 below.

4.2 College Decision

I now model the effects of skills at age 15 on a subsequent educational decision on whether to complete a college degree or not. Let D^* denote the net latent utility of completing a college education:

$$D_i^* = \alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC} + u_{iD} \tag{6}$$

Where X_{iD} is a vector of observed individual and household characteristics affecting the choice; θ^C and θ^{NC} are the unobserved abilities; u^{iD} is an error term; α_D is a vector of the coefficients associated with X_{iD}; β_D^C and β_D^{NC} indicate the effect of the corresponding factors on the decision to complete university.

D is a binary indicator that equals one if the individual completes a college degree and zero otherwise. The choice can be written as:

$$\begin{aligned} &D_i = \mathbb{1}[D_i^* > 0] \\ &or D_i = \mathbb{1}[\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC} + u_{iD} > 0] \end{aligned} \tag{7}$$

u_{iD} is assumed to be independent across the individual and household characteristics, factors and all the other errors in the model and logistically distributed. Conditional on the unobservable factor, the probability of observing D_i is:

$$Pr(D_i|X_{iD}, \theta_i) = \frac{(exp(\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC}))^{D_i}}{1 + exp(\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC})} \tag{8}$$

4.3 Labor Earnings

The model of labor earnings is given by:

$$Y_{iD} = \alpha_{Y_D} X_{iY_D} + \beta_{Y_D}^C \theta_i^C + \beta_{Y_D}^{NC} \theta_i^{NC} + u_{iY_D} \tag{9}$$

Where Y_{iD} is hourly earnings for individual i measured at age 22, $D = \{0, 1\}$ corresponding to the specific college decision above. X_{iY_D} is a vector of all other observable controls that impact earnings; θ_i^C and θ_i^{NC} are the unobserved abilities; u_{iY_D} are error terms and follow a normal distribution with mean zero and variance $\sigma^2_{u_{Y_D}}$. The probability density function of Y_{iD} is:

$$f(Y_{iD}|X_{iY_D}, \theta_i) = \frac{1}{\sqrt{2\sigma^2_{u_{Y_D}} \pi}} exp\left(-\frac{(Y_{iD} - \alpha_{Y_D} X_{iY_D} - \beta_{Y_D}^C \theta_i^C - \beta_{Y_D}^{NC} \theta_i^{NC})^2}{2\sigma^2_{u_{Y_D}}} \right) \tag{10}$$

In this model, the latent factors, θ^C and θ^{NC}, reflect unobserved heterogeneity and they are the source of dependence among observed skill measures, schooling decisions, and earnings. Controlling for these latent factors solves the problem of endogeneity arising from the endogeneity of skills and schooling and the reverse causality among skills, schooling and earnings (Heckman et al., 2006). Using latent cognitive and noncognitive skills also solves the measurement error problem. Furthermore, using skills and outcomes observed at different times deters the reverse causality among skills, schooling and labor market outcomes; skill measures are not affected by college degrees.

4.4 Estimation

Equations 3, 7 and 9 constitute the following structural model by which the college decision and wage equation are estimated jointly with the measurement system:

$$
\begin{aligned}
T_{i,math} &= \alpha_{math} + \beta^C_{math}\theta^C_i + u_{i,math} \\
T_{i,cloze} &= \alpha_{cloze} + \beta^C_{cloze}\theta^C_i + u_{i,cloze} \\
T_{i,ppvt} &= \alpha_{ppvt} + \beta^C_{ppvt}\theta^C_i + u_{i,ppvt} \\
T_{i,ser} &= \alpha_{ser} + \beta^{NC}_{ser}\theta^{NC}_i + u_{i,ser} \\
T_{i,sef} &= \alpha_{sef} + \beta^{NC}_{ses}\theta^{NC}_i + u_{i,sef} \\
T_{i,ses} &= \alpha_{ses} + \beta^{NC}_{ses}\theta^{NC}_i + u_{i,ses} \\
D_i &= \mathbb{1}[\alpha_D X_{iD} + \beta^C_D\theta^C_i + \beta^{NC}_D\theta^{NC}_i + u_{iD} > 0] \\
Y_{iD} &= \alpha_{Y_D} X_{iY_D} + \beta^C_{Y_D}\theta^C_i + \beta^{NC}_{Y_D}\theta^{NC}_i + u_{iY_D}
\end{aligned}
\tag{11}
$$

The distributions of the latent factors may follow many forms and the assumption of the factor distributions is important and must be flexible enough to capture data. I approximate the factor distributions as a mixture of two normals. This assumption ensures flexibility with fewer restrictions on the distributions (Ferguson, 1983; Attanasio et al., 2017). With this assumption, the probability density function of the factor is:

$$f(\theta) = \sum_{c=1}^{2} \tau_c f(\theta|\mu_c, \Omega_c) \tag{12}$$

Where μ_c, Ω_c and τ_c are the mean, covariance and the mixture probability of the two normals.

Let Ψ be all the parameters of the model, $\Psi = \{\alpha, \beta, \sigma, \tau_c, \mu_c, \Omega_c\}$, $\theta = \{\theta^C, \theta^{NC}\}$ be the vectors of the cognitive and noncognitive factors, $X = \{X_{i_D}, X_{iY_D}\}$. Thus, from the density and probability functions 5, 8, 10 and 12, the full model likelihood function can be derived as:

$$
\begin{aligned}
L(\Psi) &= \prod_{i=1}^{N} \iint [f(T_i|\theta^C, \theta^{NC}) \times f(Y_{i,D=1}|X_{iY_{D=1}}, \theta^C, \theta^{NC})^D \times \\
& \quad f(Y_{i,D=0}|X_{iY_{D=0}}, \theta^C, \theta^{NC})^{1-D} \times Pr(D_i|X_D, \theta^C, \theta^{NC})] dF(\theta^C) dF(\theta^{NC}) \\
&= \prod_{i=1}^{N} \iint f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta^C, \theta^{NC}) dF(\theta^C) dF(\theta^{NC}) \\
&= \prod_{i=1}^{N} \int f(T_i, D_i, Y_i|X_{i_D} X_{iY_D}, \theta) dF(\theta) \\
&= \prod_{i=1}^{N} \int f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta
\end{aligned} \tag{13}
$$

The full model log-likelihood function is

$$
\mathcal{L}(\Psi) = \sum_{i=1}^{N} ln \int f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta \tag{14}
$$

Given the unobservable nature of the factors, the likelihood function is integrated over the distributions of these unobservable factors. I estimate the log-likelihood function 14 using maximum likelihood estimation (MLE). I take a one-step estimation procedure using the minorization-maximization algorithm that is presented in Appendix C.

5 Results

5.1 Measurement System

The estimation results from the measurement system described in Equation set 3, α_j, α_k and β_j, β_k, are presented in Table 3. The measurement system examines the importance of the given latent skills, θ^C and θ^{NC}, in the six tests. The factor loadings of cognitive and noncognitive skills (β_j, β_k) on respective test scores are all significantly positive, meaning that both latent skills are positively associated with test scores as expected. Latent cognitive ability is more highly associated with the mathematics and reading comprehension (Cloze) test scores, while latent noncognitive ability more highly relates to an individual's self-esteem and self-respect and inclusion indexes. Specifically, a one standard deviation increase in cognitive ability is associated with a 0.664, 0.748 and 0.739 standard deviation increase in the PPVT, Math and Cloze scores respectively and a one standard deviation in noncognitive ability is associated with a 0.389, 0.271 and 0.492 standard deviation increase in individual self-esteem, self-efficacy and self-respect and inclusion indexes respectively.

Table 3: Measurement System

	PPVT	Math	Cloze	Self-Esteem	Self-Efficacy	Self Respect and Inclusion
Panel A: Estimated parameters						
Constant	0	-0.037**	0.009	0	-0.014	0.007
		(0.019)	(0.021)		(0.011)	(0.014)
Cognitive	1	1.126***	1.113***	-	-	-
		(0.053)	(0.055)			
Noncognitive	-	-	-	1	0.696***	1.263***
					(0.039)	(0.069)
Panel B: Average Marginal Effects of Factors (AME)[a]						
Cognitive AME	0.664***	0.748***	0.739***	-	-	-
	(0.025)	(0.019)	(0.027)			
Noncognitive AME	-	-	-	0.389***	0.271***	0.492***
				(0.015)	(0.013)	(0.017)
Average value	-0.107	-0.158	-0.110	-0.022	-0.030	-0.021
Panel C: Variance Decomposition						
Signal	0.485***	0.609***	0.513***	0.370***	0.274***	0.762***
	(0.026)	(0.022)	(0.024)	(0.022)	(0.023)	(0.039)
Noise	0.515***	0.391***	0.487***	0.630***	0.726***	0.238***
	(0.026)	(0.022)	(0.024)	(0.022)	(0.023)	(0.039)
N	738	747	739	757	757	757

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[a] Average marginal effects of a one standard deviation increase of each factor, holding other variables fixed.

To assess the information content contained in each measure from the factors and measurement errors, I calculate the contribution of each factor and measurement error in explaining the variance of the observed measures.

$$P_h^{\theta^k} = \frac{(\beta_h)^2 var(\theta_i^k)}{(\beta_h)^2 var(\theta_i^k) + var(u_{ih})} \tag{15}$$

$$P_h^{u_h^k} = \frac{var(u_{ih})}{(\beta_h)^2 var(\theta_i^k) + var(u_{ih})} \tag{16}$$

Where $P_h^{\theta^k}$ is the proportion of the variance of the hth observed measures explained by the latent factor k or signal and $P_h^{u_h^k}$ is the variance of the measure explained by the measurement error or noise. $P_h^{u_h^k}$ is the proportion of the hth measure variance that remains unexplained.

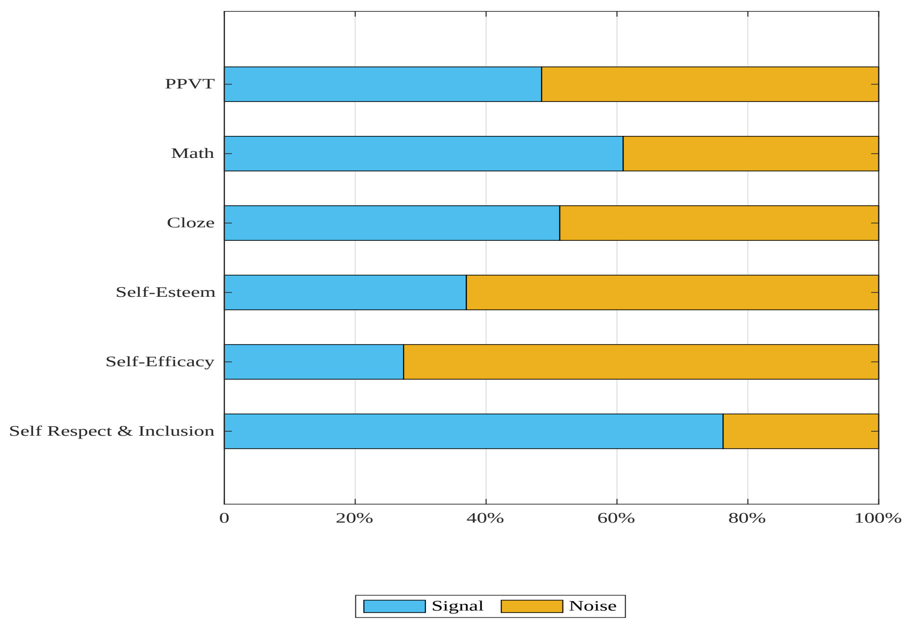
Table 3 Panel C and Figure 1 present the fraction of the variance of each measure explained by each factor (signal) and by the measurement error (noise). It is clear that the measures for each factor contain a substantial amount of information. The cognitive skill factor accounts for an important proportion of the variance of the cognitive measures - from 48.5% to 60.9%. The related measures on noncognitive skills are also informative. From 27.4% to 76.2% of the variance of the noncognitive measures are explained by signal. Although the factors explain an important proportion of the variance of the observed measures, these proportions are far from 100%, 23.8 - 72.6% of the variance of the observed measures remains unexplained and is attributed to measurement errors. This indicates that we could have serious measurement error problems if we use observed measures on their own and demonstrates the importance of the latent factor approach in measuring skills.

The estimated distributions of latent cognitive and noncognitive skills ($f(\theta^C)$, $f(\theta^{NC})$) and the skill distributions by college completion, displayed in Figures 2 and 3, show that distributions of latent skills are non-normal. These results highlight the importance of assuming a flexible distribution function for the skill distributions.[4]

Figure 3 displays the cognitive and noncognitive skill distributions by college completion. Individuals who completed college education seem to have higher cognitive and noncognitive skills; individuals with a college degree have distributions of both skills lying to the left compared to those without a college education. Although there is a difference in the

[4] The factor means, standard deviations and correlation of the estimated factors are presented in Appendix D.

Figure 1: Signals and Noises for the Measures



distributions between those with and without a college education, they show a substantial overlap. Therefore, I will also explore the effects of the variation of skills on the outcomes.

Figure 2: Estimated Cognitive Skill Distribution

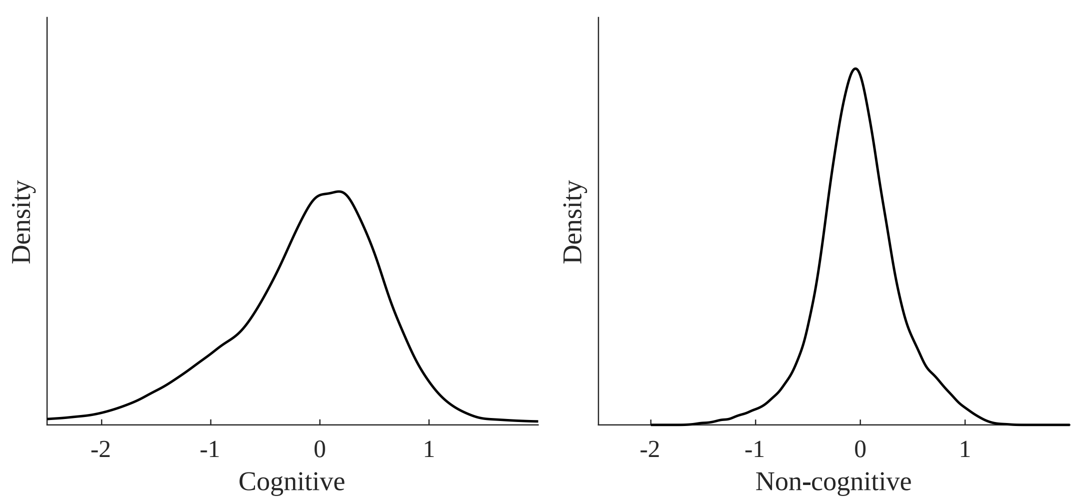
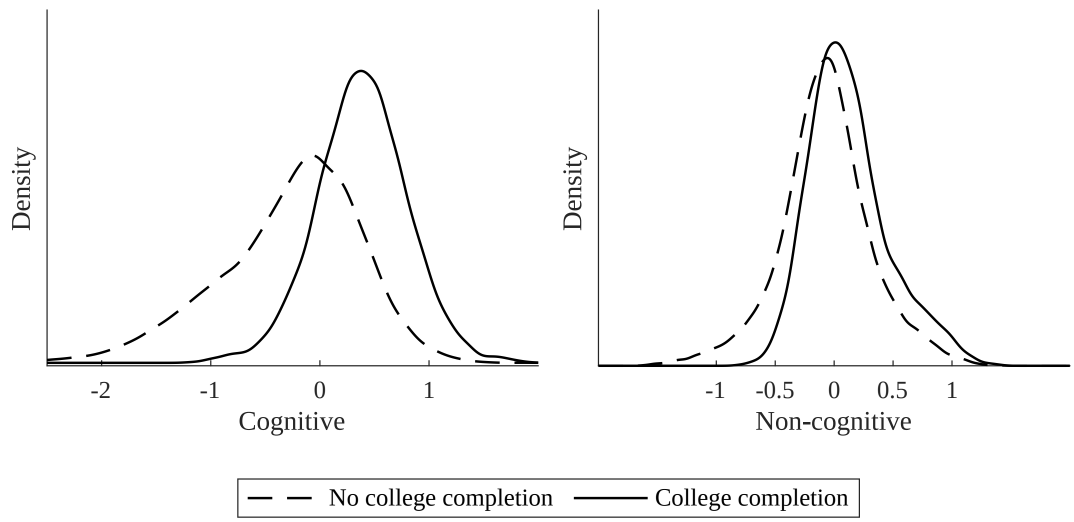


Figure 3: Estimated Skill Distribution by Educational Level



5.2 Effects of Skills on College Decision

Table 4 shows the effects of latent cognitive and noncognitive abilities on the decision to complete college. Both latent cognitive and noncognitive abilities are important determinants of completing college. The probability of completing college increases dramatically with cognitive skills and the effect of noncognitive skills is smaller. Specifically, increasing cognitive abilities by a one standard deviation would increase the likelihood of completing college by 17.2% percent and a one standard deviation increase in noncognitive abilities would lead to an increase in the likelihood of completing college by 2.6% percent.

Table 4 also shows the effects of other controlled variables on completing higher education. Girls are more likely than boys to complete college, while a child's living areas (rural/urban) do not affect the probability of completing college. Parental educational levels are important in determining college completion, while wealth does not seem to influence college completion. A youth with few siblings is more likely to obtain a college degree. The results also suggest that the child's aspiration for college education significantly determines educational attainment.

Table 4: Probability of College Completion as a Function of Skills

Variables	Coefficients	Average Marginal Effects
Cognitive	1.924***	0.172***
	(0.214)	(0.016)
Noncognitive	0.537***	0.026***
	(0.197)	(0.010)
Female	0.662***	0.082***
	(0.119)	(0.015)
Urban	0.148	0.018
	(0.120)	(0.015)
Number of siblings aged 0-18	-0.753***	-0.084***
	(0.111)	(0.010)
Wealth index	0.077	0.009
	(0.609)	(0.074)
Parental educational level	0.462***	0.059***
	(0.061)	(0.008)
Child educational aspiration	1.237***	0.146***
	(0.182)	(0.019)
Constant	-3.579***	-
	(0.415)	
Baseline probability	0.254	-
N	*757*	-

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

5.3 Effects of Skills on Earnings

Table 5 reports the estimates of the parameters, α_{Y_D}, $\beta^{C}_{Y_D}$ and $\beta^{NC}_{Y_D}$, from Equation 9 for earnings conditional on college completion. $D = 1$ and $D = 0$ indicate those who have obtained and have not obtained a college degree respectively. These results show the importance of cognitive and noncognitive skills after conditioning on college completion.

Noncognitive ability does improve income among college graduates. For this group, a one standard deviation increase in latent noncognitive ability will increase earnings by 4.107VND, which represent an substantial increase of about 17.6% over the average earnings. However, noncognitive ability does not provide any additional rewards for higher earnings among those not completing college. Cognitive ability is significantly associated with higher earnings for non-college graduates, but has no statistically significant effects on earnings for those with a college education. A one standard deviation increase in cognitive ability would increase earnings for non-college graduates by 14.7%.

These results may suggest that the effect of cognitive skills on earnings is indirect and operate mainly through educational decisions. It may be due to the fact that I examine earnings early in the career and employees are relatively new to employers. Employers had little information and opportunity to distinguish and reward higher skills, and educational levels are a meaningful signal for judgements and advancement in pay. The results also once again indicate that females earn much less than males for both groups of non-college and college graduates and individuals from urban areas earn more than those from rural areas.

Table 5: Effects of Skills on Earnings

Variables	Hourly earnings	
	D = 0	D = 1
Cognitive skills	3.338***	2.717
	(0.530)	(2.456)
Noncognitive skills	1.565	10.551***
	(1.034)	(3.081)
Cognitive AME[a]	2.217***	1.805
	(0.338)	(1.627)
Noncognitive AME[a]	0.609	4.107***
	(0.402)	(1.140)
Female	-3.822***	-2.572
	(0.524)	(1.568)
Urban	3.045***	7.526***
	(0.754)	(1.576)
Experience	2.304***	8.604***
	(0.438)	(2.418)
Experience squared	-0.309***	-1.724**
	(0.056)	(0.698)
Constant	14.726***	13.885***
	(0.662)	(1.676)
Average value	15.116	23.282

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.
[a] Average Marginal Effects of Factors.

One important advantage of structural models is the ability to simulate counterfactual outcomes (Heckman et al., 2011; Carneiro et al., 2003). To understand the effect of having a college degree, I calculate the average treatment effect (ATE) of a college degree, the treatment effect on the treated (TOT) and the treatment effect on the untreated (TOU) as follows:

$$\begin{aligned} ATE &= E[Y_1 - Y_0|X,\theta] = E[Y_1|X,\theta) - (Y_0|X,\theta)] \\ TOT &= E[Y_1 - Y_0|X,\theta,D=1] = E[Y_1|X,\theta,D=1) - (Y_0|X,\theta,D=1)] \\ TOU &= E[Y_0 - Y_1|X,\theta,D=0] = E[Y_0|X,\theta,D=0) - (Y_1|X,\theta,D=0)] \end{aligned} \tag{17}$$

Table 6 shows the difference between the means of earnings conditioning on the decision to complete a college degree and the respective counterfactual earnings. ATE and TOT are positive and TOU is negative. The results suggest that, on average, young people with a college degree would have higher earnings. Even people with their given background and latent skills who decided not to have a college degree would have higher earnings if they had a college degree. In particular, on average, individuals would increase their hourly earnings by 3,082 (equivalent to 17.9% relative to the mean earnings) if they decided to have a college degree. Conditioning on completing a college degree, the mean of hourly earnings is 6,536 (equivalent to 38% relative to the mean earnings) higher than the mean of hourly earnings that they would have earned if they had decided not to have a college degree. In contrast, the mean of hourly earnings conditional on not completing a college degree is 1,909 (equivalent to 11.1% relative to the mean earnings) lower than the means of counterfactual hourly earnings that they would have earned if they had decided to complete a college degree.

Figures 4, 5 and 6 graphically present how the college decision and earnings vary across deciles of cognitive and noncognitive abilities. In these figures, I present each outcome as a function of deciles of the skill distribution and display the mean value of these outcomes by deciles of the skills.

Figure 4 shows the probability of college completion by each decile of the cognitive and noncognitive skill distribution. Both types of skills show strong effects on the probability of completing a college degree. A steeper gradient for cognitive ability shows that its effect on

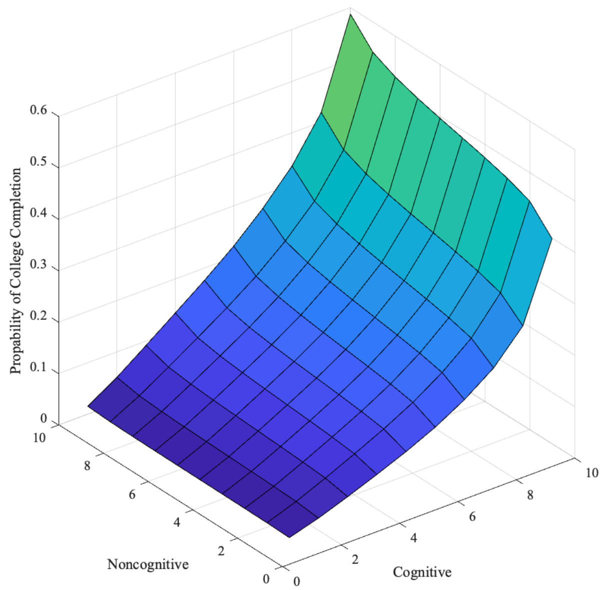
Table 6: Treatment Effects

	Estimates
$E[Y_1\|X,\theta) - (Y_0\|X,\theta)]$	3.082*
	(2.200)
$E[Y_1\|X,\theta,D=1) - (Y_0\|X,\theta,D=1)]$	6.536***
	(0.721)
$E[Y_0\|X,\theta,D=0) - (Y_1\|X,\theta,D=0)]$	-1.909
	(2.893)

Notes: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

college completion is more important than noncognitive ability. The probability of graduating from a college increases dramatically with cognitive skills while the effect of noncognitive skills is stronger for a higher level of cognitive skills.

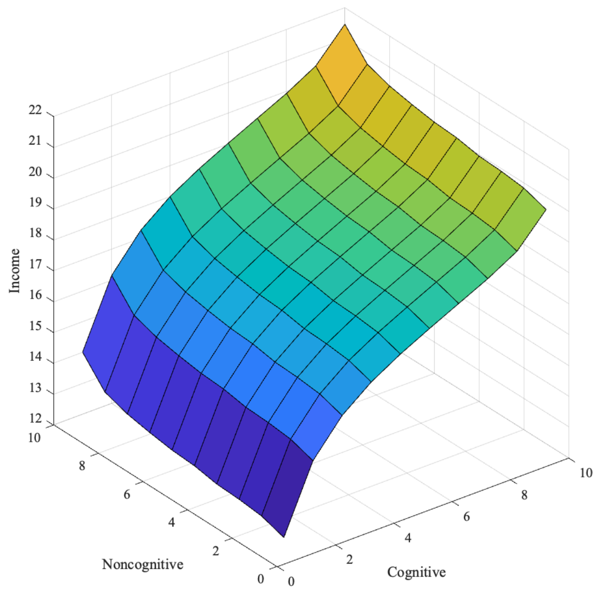
Figure 4: Probability of College Completion by Deciles of the Skills



Note: z-axis is the probability of college completion within pairs of deciles of the cognitive and noncognitive factors, x-axis and y-axis are deciles of the cognitive and noncognitive factors respectively.

Figure 5 displays the effects of skills on earnings by deciles of the skill distribution. The effect of cognitive skills is again stronger than noncognitive skills. This result can be explained by the fact that the skills not only have direct effects on earnings, but also have indirect effects on earnings through schooling that generates effects on earnings, while the effect of the cognitive skills on schooling is more important than noncognitive skills.

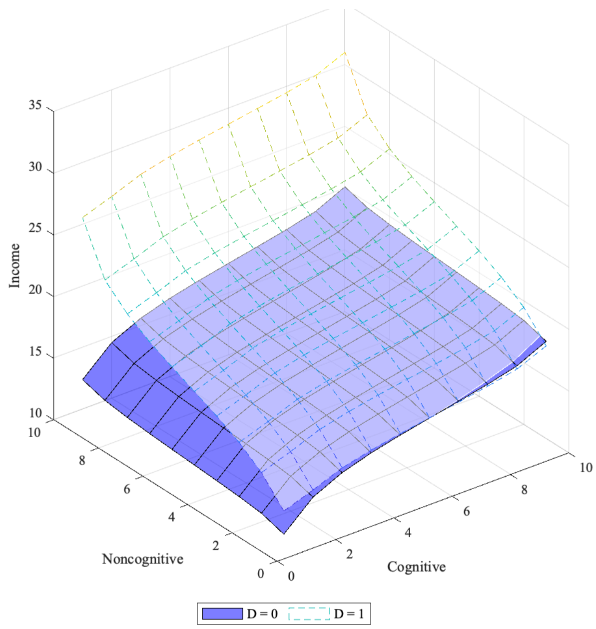
Figure 5: Earnings by Deciles of the Skills



Note: z-axis is the mean earnings within pairs of deciles of the cognitive and noncognitive factors, x-axis and y-axis are deciles of the cognitive and noncognitive factors respectively.

Figure 6 shows the effects of skills on earnings by college completion across deciles of skills. Earnings increase across deciles of cognitive and noncognitive abilities from about 10,000 to 20,000VND and from 12,000 to 31,000VND for those without and with a college degree respectively. Noncognitive skills play a significant role in earnings for those with a college degree, while those without a college degree need a certain level of cognitive skills to get higher earnings.

Figure 6: Earnings by College Completion by Deciles of the Skills



Note: z-axis is the mean earnings within pairs of deciles of the cognitive and noncognitive factors, x-axis and y-axis are deciles of the cognitive and noncognitive factors respectively. D = 1: college completion, D = 0 otherwise.

The results share commonalities with the literature. First, both cognitive and noncognitive abilities affect schooling and labor market earnings (Almlund et al., 2011; Hanushek and Woessmann, 2008; Almlund et al., 2011; Hanushek, 2009; Heckman et al., 2006). Second, the effects of these abilities on earnings are mediated by levels of schooling (Heckman et al., 2006; Heckman et al., 2011). Third, our results about the effects of skills on schooling are consistent with Cunha et al. (2010) and Duncan et al. (2007) in the sense that both cognitive and noncognitive skills have effects on schooling, while cognitive skills have stronger effects.

Despite these consistencies, there are certain differences in terms of effects and magnitudes of effects of skills on education decisions and labor market earnings between my findings and the literature. These deferences arise from different reasons. First, the differences in markets, policies and institutions can result in different labor outcomes. Second, different questionnaires, measures to capture the underlying cognitive and noncognitive skills and different methodologies may influence the conclusions about the effects. My findings indicate that noncognitive skills are highly valued, while cognitive skills are not rewarded once students graduate from college. This result is in contrast with Heckman et al. (2006) for the US who found that noncognitive traits have little value, while cognitive skills have a strong effect on earnings for 4-year-college graduates and both skills have strong effects in the 2-year-college market. This reflects the fact in Vietnam that the education system equipped the workforce with a low level of 'soft' skills and there is a high demand for these skills (Bodewig et al., 2014).

Appendix E shows the results for an alternative specification for the measurement system that allows for correlated cognitive and noncognitive factors where each of the cognitive measures depends on both the cognitive and noncognitive factors and the noncognitive measures are a function of the noncognitive factor only. I impose the same normalizations on the scales and locations of the factors. The results in Appendix Tables E.2 and E.3 show few differences in the effects of skills on college completion and earnings between the two specifications. Although noncognitive abilities play a smaller role, but the difference is insignificant.

6 Conclusion

This study uses high-quality data from the Vietnam Young Lives survey, Older Cohorts and the latent factor approach with a two-dimensional latent factor structure to examine the roles of both cognitive and noncognitive abilities in explaining schooling decisions and subsequent earnings in Vietnam. The results suggest that both cognitive and noncognitive skills play a role in determining earnings. The analysis shows that the effects of skills on earnings operate not only indirectly through the educational channel but also directly in the labor market. Because of the nature of endogenous schooling decisions, the dynamics in decision making is crucial in investigating the effects of skills on earnings. Among college graduates, noncognitive skills not only directly influence earnings in the labor market but also have indirect effects through educational choices. The results suggest that it is equally important to improve noncognitive skills as cognitive skills.

There is strong evidence showing that human capital is shaped early in the life cycle and skills beget skills in a complementary and dynamic fashion. Child development at an early age has direct long-lasting effects on social and economic outcomes for individuals and society. Policies should give equal attention to improving different dimensions of noncognitive skills in early childhood as with cognitive skills. This is especially true for Vietnam, which achieves impressive results in cognitive skills, but soft skills and labor productivity are relatively low. Skills are affected by a combination of inputs, including individual abilities, family investments, and home, school, and community environments. Therefore, policies should consider a combination of factors, and investments in childhood development are a cost-effective strategy for improving productivity, promoting economic growth, and reducing inequality.

References

Acosta, P., N. Muller, and M. A. Sarzosa (2015). *Beyond qualifications: returns to cognitive and socio-emotional skills in Colombia*. The World Bank.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality psychology and economics. Volume 4, Chapter 1, pp. 1–181. Elsevier.

Attanasio, O., C. Meghir, E. Nix, and F. Salvati (2017). Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of Economic Dynamics 25*, 234–259. Special Issue on Human Capital and Inequality.

Aucejo, E. and J. James (2021). The path to college education: The role of math and verbal skills. *Journal of Political Economy 129*(10), 2905–2946.

Bodewig, C., R. Badiani-Magnusson, K. Macdonald, D. Newhouse, and J. Rutkowski (2014). *Skilling up Vietnam: Preparing the workforce for a modern market economy*. The World Bank.

Borghans, L., A. L. Duckworth, J. Heckman, and B. ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4).

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276. PMID: 26828106.

Cawley, J., J. Heckman, and E. Vytlacil (2001). Three observations on wages and measured cognitive ability. *Labour Economics 8*(4), 419 – 442.

Crookston, B., R. Forste, C. McClellan, A. Georgiadis, and T. Heaton (2014). Factors associated with cognitive achievement in late childhood and adolescence: The young lives cohort study of children in ethiopia, india, peru, and vietnam. *BMC pediatrics 14*, 253.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources 43*(4), 738–782.

Cunha, F. and J. J. Heckman (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association 7*(2-3), 320–364.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cunningham, W., M. P. Torrado, and M. Sarzosa (2016). *Cognitive and non-cognitive skills for the Peruvian labor market: Addressing measurement error through latent skills estimations*. The World Bank.

Dercon, S. and P. Krishnan (2009). Poverty and the psychosocial competencies of children: evidence from the young lives sample in four developing countries. *Children Youth and Environments 19*(2), 138–163.

Dercon, S. and A. Sánchez (2013). Height in mid childhood and psychosocial competencies in late childhood: Evidence from four developing countries. *Economics and Human Biology 11*(4), 426 – 432.

Díaz, J. J., O. Arias, and D. V. Tudela (2012). Does perseverance pay as much as being smart? the returns to cognitive and non-cognitive skills in urban Peru. *Unpublished paper, World Bank, Washington, DC*.

Drago, F. (2011). Self-esteem and earnings. *Journal of Economic Psychology 32*(3), 480 – 488.

Duncan, G., C. Dowsett, A. Claessens, K. Magnuson, A. Huston, P. Klebanov, L. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, H. Sexton, and C. Japel (2007). School readiness and later achievement. *Developmental psychology 43*, 1428–46.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, pp. 287–302. Elsevier.

GSO (2016). Nang suat lao dong Viet Nam: Thuc trang va giai phap [Vietnam's labor productivity: Situations and solutions].

Glewwe, P., Q. Huang, and A. Park (2017). Cognitive skills, noncognitive skills, and school-to-work transitions in rural China. *Journal of Economic Behavior & Organization 134*, 141 – 164.

Green, D. and W. C. Riddell (2003). Literacy and earnings: An investigation of the interaction of cognitive and unobserved skills in earnings generation. *Labour Economics 10*, 165–184.

Hanushek, E. A. (2002). Publicly provided education. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Volume 4 of *Handbook of Public Economics*, Chapter 30, pp. 2045–2141. Elsevier.

Hanushek, E. A. (2009). *The economic value of education and cognitive skills*, Chapter 3. Routledge.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to skills around the world: Evidence from piaac. *European Economic Review 73*, 103 – 130.

Hanushek, E. A. and L. Woessmann (2008). The role of cognitive skills in economic development. *Journal of Economic Literature 46*(3), 607–68.

Hanushek, E. A. and L. Zhang (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital 3*(2), 107–143.

Heckman, J., J. Humphries, S. Urzúa, and G. Veramendi (2011). The effects of educational choices on labor market, health, and social outcomes. *Human Capital and Economic Opportunity Working Paper*.

Heckman, J. J., J. Stixrud, and S. Urzúa (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

James, J. (2017). MM algorithm for general mixed multinomial logit models. *Journal of Applied Econometrics 32*(4), 841–857.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*(1), 141–151.

Kottelenberg, M. J. and S. F. Lehrer (2019). How skills and parental valuation of education influence human capital acquisition and early labor market return to human capital in canada. *Journal of Labor Economics 37*(S2), S735–S778.

Krishnan, P. and S. Krutikova (2013). Non-cognitive skill formation in poor neighbourhoods of urban India. *Labour Economics 24*, 68 – 85.

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review 10*(2), 179–214.

Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics 3*(1), 101–28.

Long, M. C., D. Goldhaber, and N. Huntington-Klein (2015). Do completed college majors respond to changes in wages? *Economics of Education Review 49*, 1 – 14.

Murnane, R. J., J. B. Willett, M. Braatz, and Y. Duhaldeborde (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? evidence from the NLSY. *Economics of Education Review 20*(4), 311 – 320.

Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management 19*(4), 547–568.

Nguyen, N. (2008, 01). An assessment of the young lives sampling approach in vietnam. *University of Oxford, Open Access publications from University of Oxford*.

Nordman, C. J., L. Sarr, and S. Sharma (2015). Cognitive, non-Cognitive skills and gender wage gaps: Evidence from linked employer-employee data in Bangladesh. IZA Discussion Papers 9132, Institute of Labor Economics (IZA).

Prada, M. F. and S. Urzúa (2017). One size does not fit all: Multiple dimensions of ability, college attendance, and earnings. *Journal of Labor Economics 35*(4), 953–991.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton University Press.

Roseth, V. V., A. Valerio, and M. Gutierrez (2016). *Education, skills, and labor market outcomes: Results from Large-Scale Adult Skills Surveys in urban areas in 12 countries.* World Bank.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied 80*(1), 1.

Sahn, D. E. and K. M. Villa (2015). The role of personality, cognition and shocks in determining age of entry into labor market, sector of employment, and within sector earnings. (330-2016-13805), 83.

Sánchez, A. (2013). The structural relationship between nutrition, cognitive and noncognitive skills: evidence from four developing countries. Young Lives Working Paper 111.

Sánchez, A. and A. Singh (2018). Accessing higher education in developing countries: Panel data analysis from India, Peru, and Vietnam. *World Development 109*, 261 – 278.

Singh, A. (2019). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*. jvz033.

Thiel, H. and S. L. Thomsen (2013). Noncognitive skills in economics: Models, measurement, and empirical evidence. *Research in Economics 67*(2), 189 – 214.

Urzúa, S. (2008). Racial labor marketgaps: The role of abilities and schooling choices. *The Journal of Human Resources 43*(4), 919–971.

VandenBos, G. R. (2007). *APA dictionary of psychology.* American Psychological Association.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted jamaican children: prospective cohort study. *The lancet 366*(9499), 1804–1807.

Appendix A: Description of Variable Construction

Table A.1: Description of Variable Construction

Variables	Description
Cognitive Skills	
PPVT score	The PPVT is a test of receptive vocabulary. It uses a stimulus word and accompanying pictures to test receptive vocabulary. The PPVT-III with 204 items is used in the Young Lives survey in Vietnam. PPVT scores are standardized scores in The PPVT.
Math score	The mathematics test (Math test) include 29 items on addition, subtraction, multiplication, division, problem-solving, mea- surement, data interpretation, and basic geometry. Math scores are standardized scores in math test.
Cloze test scores	The Cloze test is developed to measure verbal skills and reading comprehension. The test include 24 items that increase in difficulty. Cloze test scores are standardized scores in Cloze test.
Noncognitive Skills	
Self-esteem	The self-esteem scale is constructed as the average of the following standardized items/statements (five-point Likert scales). 1. 'I am proud of my clothes'; 2. 'I feel my clothing is right for all occasions'; 3. 'I am proud of my shoes or of having shoes'; 4. 'I am proud because I have the right books, pencils or other equipment for school'; 5. 'I am proud that I have the correct uniform'; 6. 'I am proud of the work I have to do'.

Continued on next page

Table A.1: Description of Variable Construction *Continued*

Variables	**Description**
Self-efficacy	The Self-efficacy index is the average of the following standardized items (five-point Likert scales): 1. ‘If we try hard we can improve my situation in life’; 2. ‘Other people in my family make all the decisions about how we spend my time’; 3. ‘I like to make plans for my future studies and work’; 4. ’If we study hard we will be rewarded with a better job in the future’; 5. ‘I have no choice about the work I do - I must do this sort of work’.
Self-respect and Inclusion	This index is the average of the following standardized items (five-point Likert scales): 1. ‘When I am at the shops/market I am usually treated by others with fairness and respect’; 2. ‘Adults in my community treat me as well as they treat other children of my age’; 3. ‘The other children in my class treat me with respect’; 4. ‘Other pupils in my class tease me at school’; 5. ‘My friends will stand by me during difficult times’; 6. ‘I feel I belong at my school’; 7. ‘My friends look up to me as a leader’; 8. ‘I have people I look up to’ 9. ‘I have opportunities to develop job skills’.
Other Variables	
Hourly earnings	Hourly earnings from all paid activities by child in the past 12 months.

Continued on next page

Table A.1: Description of Variable Construction *Continued*

Variables	**Description**
College completion	Binary variable equal to one if the individual completes a college degree and zero otherwise.
Wealth Index	The wealth index is a composite measure of living standards, it is the average of the three sub-indexes: consumer durable, housing quality and access to service indexes. It takes values from 0 to 1, a higher value reflect a wealthier household.
Parental educational level	The highest level of education of Parent: 1 = less than primary; 2 = primary; 3 = Lower secondary; 4 = Upper secondary; 5 = post-seconday.

Appendix B: Exploratory Factor Analysis

This appendix provides details of the factor analysis to find whether there are factors that represent cognitive skills and noncognitive skills and how many factors retained.

Kaiser's eigenvalue rule: The Kaiser's criterion consists of retaining only factors with eigenvalues greater than 1 (Kaiser, 1960). The intuition behind this rule is that a factor should extract more variance than contained in a single variable, otherwise it should be dropped.

Scree plot: The scree plot was introduced by Cattell (1966). It is a visual tool used to help determine the number of important factors based on the analyst's inspection of a plot of the eigenvalues associated with the data. The number of factors should be equal to the number of eigenvalues before which the smooth decrease of eigenvalues levels off to the right of the plot.

Cognitive skills:

Table B.1: Factor Analysis/Correlation - Cognitive Skills (Principal Component Factors)

	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.99155	1.40944	0.66385	0.66385
Factor2	0.58211	0.15577	0.19404	0.85789
Factor3	0.42634	.	0.14211	1.00000

LR test: independent vs. saturated: chi2(3) = 504.68.

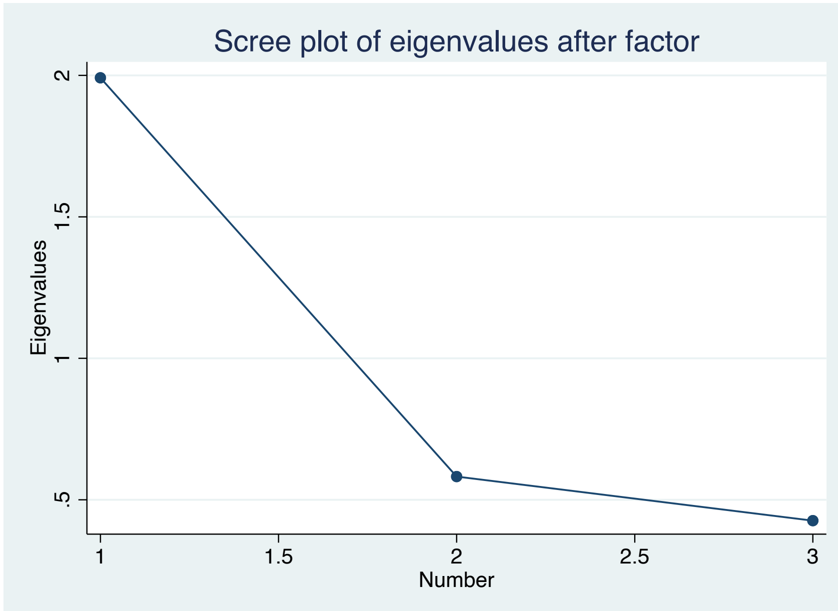
Prob>chi2 = 0.0000.

Retained factors = 1, 718 observations.

Table B.2: Factor Loadings (Pattern Matrix) and Unique Variances - Cognitive Skills

	Factor1	Uniqueness
PPVT test	.7873078	.3801465
Math test	.8538751	.2708974
Cloze test	.8016167	.3574107

Figure B.1: Scree Plot - Cognitive Skills



Noncognitive Skills:

Table B.3: Factor Analysis/Correlation - Noncognitive Skills (Principal Component Factors)

	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.88172	1.20759	0.62724	0.62724
Factor2	0.67413	0.22998	0.22471	0.85195
Factor3	0.44415	.	0.14805	1.00000

LR test: independent vs. saturated: chi2(3) = 433.27.

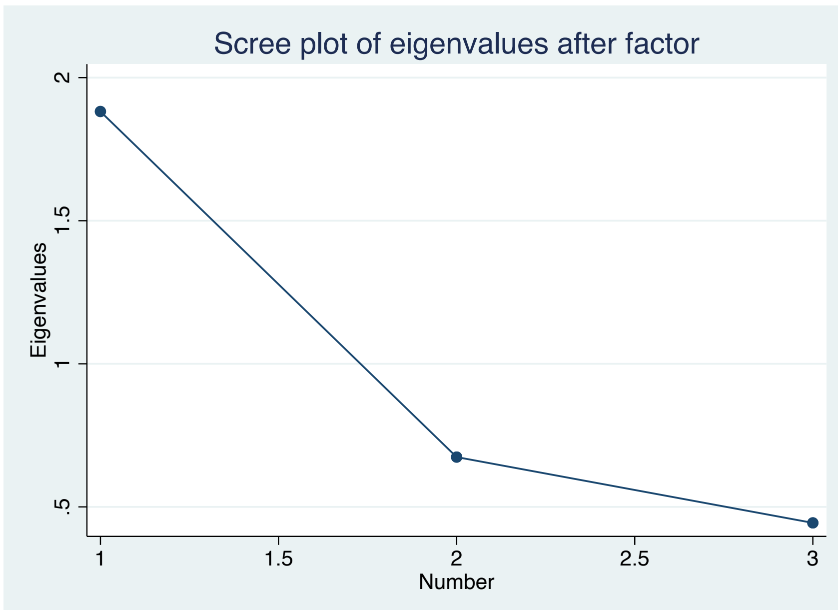
Prob>chi2 = 0.0000.

Retained factors = 1, 757 observations.

Table B.4: Factor Loadings (Pattern Matrix) and Unique Variances - Noncognitive Skills

	Factor1	Uniqueness
Self-esteem	.7866501	.3811816
Self-efficacy	.7349131	.4599027
Self Respect and Inclusion	.8501783	.2771969

Figure B.2: Scree Plot - Noncognitive Skills



Appendix C: Model Estimation Procedure

The full log-likelihood function I want to estimate is Equation 14:

$$\mathcal{L}(\Psi) = \sum_{i=1}^{N} ln\left(\int f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta\right) \tag{C.1}$$

Where Ψ are all the parameters of the model that I want to estimate, $\Psi = \{\alpha, \beta, \sigma, \tau_c, \mu_c, \Omega_c\}$.

I maximize the log-likelihood function C.1 using the minorization-maximization algorithm developed in James (2017) and Aucejo and James (2021).

Given an initial value of parameters, Ψ^0, the log-likelihood function $\mathcal{L}(\Psi)$ can be bounded below by a quadratic function:

$$Q(\Psi|\Psi^0) = \sum_{i=1}^{n} \int ln(f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta) f(\theta)) h(\theta | T_i, D_i, Y_i, \Psi^0) d\theta \tag{C.2}$$

where

$$h(\theta | T_i, D_i, Y_i, \Psi^0) = \frac{f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta) f(\theta)}{\displaystyle\int f(T_i, D_i, Y_i | X_{i_D}, X_{iY_D}, \theta') f(\theta') d\theta'} \tag{C.3}$$

Given the integral in the surrogate function $Q(\Psi|\Psi^0)$, it must be simulated by drawing R values of θ from $f(\theta|\Psi^0)$ and approximating $h(\theta|B_i, P_i, \Psi^0)$ by the weight:

$$w_{ir}^0 = \frac{f(T_i, D_i, Y_i | X_{ir_D}, X_{irY_D}, \theta_{ir}^0)}{\displaystyle\sum_{r=1}^{R} f(T_i, D_i, Y_i | X_{ir_D}, X_{irY_D}, \theta_{ir}^0)} \tag{C.4}$$

The lower bound function is now:

$$Q(\Psi|\Psi^0) = \sum_{i=1}^{n} \sum_{r=1}^{R} w_{ir}^0 ln(f(T_i, D_i, Y_i | X_{ir_D}, X_{irY_D}, \theta_{ir}^0) f(\theta_{ir}^0)) \tag{C.5}$$

Maximizing this function gives a new set of parameters, Ψ^1, that guarantee $\mathcal{L}(\Psi^1) > \mathcal{L}(\Psi^0)$. Replacing Ψ^1 with Ψ^0 and iterating this process until the parameters converge. Let m denote the mth iteration of the algorithm. The parameter updates at the mth iteration are found by:

$$\hat{\tau}_{ir} = w_{ir}^{m} \frac{\tau_c^m normpdf(\theta_{ir}^m, \mu_c^m, \Omega_c^m)}{\sum_{c'=1}^{C} \tau_{c'}^m normpdf(\theta_{ir}^m, \mu_{c'}^m, \Omega_{c'}^m)} \tag{C.6}$$

$$\begin{aligned}
\tau_c^{m+1} &= \frac{\sum_{i=1}^{N}\sum_{i=1}^{R} \hat{\tau}_{ir}}{n} \\
\mu_c^{m+1} &= \frac{\sum_{i=1}^{N}\sum_{i=1}^{R} \hat{\tau}_{ir}\theta_{ir}^m}{\sum_{i=1}^{N}\sum_{i=1}^{R} \hat{\tau}_{ir}} \\
\Omega_c^{m+1} &= \frac{\sum_{i=1}^{n}\sum_{i=1}^{R} \hat{\tau}_{ir}(\theta_{ir}^m)(\theta_{ir}^m)'}{\sum_{i=1}^{n}\sum_{i=1}^{R} \hat{\tau}_{ir}} - (\mu_c^{m+1})(\mu_c^{m+1})'
\end{aligned} \tag{C.7}$$

Since θ are treated as observed variables, the updated parameters $\{\alpha, \beta, \sigma\}$ can be estimated by standard OLS and logit models for the continuous and binary dependent variables respectively with the weights. In particular, for simplicity, let y_i be dependent variables including the observed measures, college choice and income and x_i be independent variables, including observed covariates and unobserved factors. Equation system 11, which I want to estimate, take the form $y_i = {x_i}'\beta + u_i$.

If y_i is continuous, then

$$\begin{aligned}
&\beta^{m+1} = (XX)^{-1} * XY \\
&Where\ XX = \sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^m (x_{ir})(x_{ir})' \\
&and\ XY = \textstyle\sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^m (x_{ir})(y_i)
\end{aligned} \tag{C.8}$$

If y_i is binary, then

$$\begin{aligned}
&\beta^{m+1} = \beta^m - B^{-1} * XY \\
&Where\ B = -\tfrac{1}{4}\sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^m (x_{ir})(x_{ir})' \\
&and\ XY = \sum_{i=1}^{N}\sum_{i=1}^{R} w_{ir}^m (x_{ir})({y_i}' - p_{ir}^{m\prime}) \\
&with\ p_{ir}^m = \frac{(exp({x_{ir}}'\beta))^{D_i}}{1 + exp({x_{ir}}'\beta)}
\end{aligned} \tag{C.9}$$

Appendix D: Factor Distribution Moments

Table D.1: Factor Means, Standard Deviation and Correlation

	Cognitive skills	Noncognitive skills
Factore means	-0.107	-0.022
	(0.020)	(0.011)
Factor standard deviation	0.664	0.389
	(0.025)	(0.015)
Factor correlation:		
Cocnitiveskill	1	–
Noncognitiveskill	0.290	1
	(0.028)	

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Table D.2: Mixture Component Means

	Cognitive skills	Noncognitive skills	Type share
Type 1	-0.351	-0.011	0.516
	(0.098)	(0.034)	(0.071)
Type 2	0.152	-0.034	0.484
	(0.041)	(0.034)	(0.071)

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process.

Appendix E: An Alternative Specification for the Factors

An alternative setting to the factor loadings is triangular as follows:

$$T_{ij} = \alpha_j + \beta_j^C \theta_i^C + \beta_j^{NC} \theta_i^{NC} + u_{ij} \tag{E.1}$$

for j = {1, 2, 3}= {math, cloze, ppvt}.

$$T_{ik} = \alpha_k + \beta_k \theta_i^{NC} + u_{ik} \tag{E.2}$$

for k = {1, 2, 3}= {ses, sef, ser}.

Specifically, the measurement system takes the following form:

$$\begin{aligned}
T_{i,ppvt} &= \alpha_{ppvt} + \beta_{ppvt}^C * \theta_i^C + \beta_{ppvt}^{NC} \theta_i^{NC} + u_{i,ppvt} \\
T_{i,math} &= \alpha_{math} + \beta_{math}^C \theta_i^C + \beta_{math}^{NC} \theta_i^{NC} + u_{i,math} \\
T_{i,cloze} &= \alpha_{cloze} + \beta_{cloze}^C \theta_i^C + \beta_{cloze}^{NC} \theta_i^{NC} + u_{i,cloze} \\
T_{i,ses} &= \alpha_{ses} + \beta_{ses}^{NC} * \theta_i^{NC} + u_{i,ses} \\
T_{i,sef} &= \alpha_{sef} + \beta_{sef}^{NC} \theta_i^{NC} + u_{i,sef} \\
T_{i,ser} &= \alpha_{ser} + \beta_{ser}^{NC} \theta_i^{NC} + u_{i,ser}
\end{aligned} \tag{E.3}$$

The factor loadings are as follows:

$$\left[\beta_T^C, \beta_T^{NC}\right] = \begin{bmatrix} \beta_{math}^C & \beta_{math}^{NC} \\ \beta_{cloze}^C & \beta_{cloze}^{NC} \\ \beta_{ppvt}^C & \beta_{ppvt}^{NC} \\ \beta_{ser}^C & \beta_{ser}^{NC} \\ \beta_{sef}^C & \beta_{sef}^{NC} \\ \beta_{ses}^C & \beta_{ses}^{NC} \end{bmatrix} = \begin{bmatrix} \beta_{math}^C & \beta_{math}^{NC} \\ \beta_{cloze}^C & \beta_{cloze}^{NC} \\ 1 & \beta_{ppvt}^{NC} \\ 0 & \beta_{ser}^{NC} \\ 0 & \beta_{sef}^{NC} \\ 0 & 1 \end{bmatrix} \tag{E.4}$$

Where both the cognitive and noncognitive factors load onto or affect the observed cognitive measures and only the noncognitive factor load onto the noncognitive measures.

All of the equations and the analysis on the schooling decision and earnings outcome are the same as in the main text.

Tables E.1 , E.2 and E.3 show the estimates of this alternative setting. The coefficients of controls, loadings and latent factors are not much different from the main specification.

Table E.1: Measurement System - Correlated Factors

	PPVT	Math	Cloze	Self-Esteem	Self-Efficacy	Self Respect and Inclusion
Panel A: Estimated parameters						
Constant	0	-0.038**	0.010	0	-0.014	0.008
		(0.018)	(0.022)		(0.011)	(0.014)
Cognitive	1	1.140***	1.131***	0	0	0
		(0.058)	(0.060)			
Noncognitive	0	-0.063	-0.021	1	0.707***	1.280***
		(0.069)	(0.076)		(0.040)	(0.074)
Panel B: Average Marginal Effects of Factors (AME)[a]						
Cognitive AME	0.661***	0.754***	0.747***	0	0	0
	(0.025)	(0.022)	(0.028)			
Noncognitive AME	0	-0.024	-0.008	0.385***	0.272***	0.493***
		(0.026)	(0.029)	(0.016)	(0.013)	(0.016)
Average value	-0.106	-0.157	-0.109	-0.023	-0.030	-0.021
Panel C: Variance Decomposition						
Signal	0.479***	0.608***	0.521***	0.363***	0.276***	0.765***
	(0.026)	(0.023)	(0.022)	(0.024)	(0.023)	(0.041)
Noise	0.521***	0.392***	0.479***	0.637***	0.724***	0.235***
	(0.026)	(0.023)	(0.022)	(0.024)	(0.023)	(0.041)
N	738	747	739	757	757	757

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

[a] Average Marginal Effects of Factors.

Table E.2: Probability of College Completion as a Function of Skills - Correlated Factors

Variables	Coefficients	Average Marginal Effects
Cognitive	1.949***	0.174***
	(0.221)	(0.017)
Noncognitive	0.473**	0.023**
	(0.196)	(0.010)
Female	0.658***	0.082***
	(0.119)	(0.015)
Urban	0.154	0.019
	(0.121)	(0.015)
Number of siblings aged 0-18	-0.748***	-0.084***
	(0.111)	(0.010)
Wealth index	0.080	0.010
	(0.608)	(0.074)
Parental educational level	0.463***	0.060***
	(0.061)	(0.008)
Child's educational aspiration	1.220***	0.145***
	(0.180)	(0.019)
Constant	-3.580***	-
	(0.418)	
Baseline probability	0.254	-
N	*757*	-

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Table E.3: Effects of Skills on Earnings - Correlated Factors

Variables	Hourly earnings	
	D = 0	D = 1
Cognitive	3.299***	3.131
	(0.538)	(2.410)
Noncognitive	1.425	10.470***
	(1.022)	(2.982)
Cognitive AME	2.180***	2.069
	(0.339)	(1.590)
Noncognitive AME	0.549	4.030***
	(0.397)	(1.097)
Female	-3.818***	-2.554
	(0.524)	(1.570)
Urban	3.079***	7.578***
	(0.750)	(1.571)
Experience	2.300***	8.527***
	(0.441)	(2.411)
Experience squared	-0.308***	-1.713**
	(0.056)	(0.692)
Constant	14.695***	13.767***
	(0.663)	(1.673)
Average value	15.114	23.287

Note: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.
[a] Average Marginal Effects of Factors.