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

Duc Thanh Nguyen

April 20, 2026

Motivation

- Adolescence is a critical development period characterized by biological, cognitive and social-emotional changes.
- These changes can profoundly impact developmental trajectories in emerging adulthood and over the life course.
- In this period, youth is also susceptible to surrounding environments.
- Children in developing countries face various risk factors and deficits in human capital development.
- Roughly one in three students are regularly bullied.

Objectives

- Model latent/true cogntive and noncognitive skills to capture the dimensions of skills more accurately.
- Assess the effects of these skills on endogenous educational choices and subsequent earnings.

Main contributions

Extend the literature on the effects of different skills, especially noncognitive skills, on college completion choices and labor market outcomes in developing countries:

- use a structural latent factor approach to identify latent skills to capture multiple skill dimensions and correct any measurement errors.
- explicitly embed a model of endogenous education decisions and subsequent earnings into a latent factor model to solve the endogeneity problems and identify the channels through which skills affects schooling and labor market earnings.

Literature review

- Use test scores as a proxy for cognitive and noncognitive skills.
- Focus more on cognitive skills and developed countries.
- Suffer from the problems of endogeneity.
- Developed countries: Heckman et al. (2006), Cunha et al. (2010), and Heckman et al. (2011) developed and used structural measurement frameworks to address measurement errors in measuring skills and the endogeneity of observed skills, schooling and outcomes.
- Developing countries: studies suffer from either measurement error or the endogeneity problems.

Data: Young Lives survey (YL)

- Older Cohort: followed 1,000 children every 4 years from age 8 to 22.
- YL provides a rich dataset on diversified aspects of children, their families and communities.
- Cognitive skills:
  - Mathematics test, Peabody Picture Vocabulary Test (PPVT) and sentence comprehension test (Cloze test).
- Noncognitive skills:
  - Self-esteem: overall evaluation of their worth.
  - Self-efficacy: sense of agency or mastery over life.
  - Self-respect and inclusion: sense of pride and inclusion.

Model structure

- Follows Heckman et al. (2006), Carneiro et al. (2003), Cunha et al. (2010).
- Models observed test scores and noncognitive skill measures as a function of latent/true cognitive and noncognitive skills.
- Educational investment decision (college completion) is endogenous, determined by prevously accumulated latent skills and might affect future earnings.
- Investment-dependent earnings.

Model structure

- Measurement system:

$$
\begin{aligned}
T_{i,math} &= \alpha_{math} + \beta^{C}_{math}\theta^{C}_{i} + u_{i,math} \\
T_{i,cloze} &= \alpha_{cloze} + \beta^{C}_{cloze}\theta^{C}_{i} + u_{i,cloze} \\
T_{i,ppvt} &= \alpha_{ppvt} + \beta^{C}_{ppvt}\theta^{C}_{i} + u_{i,ppvt} \\
T_{i,ser} &= \alpha_{ser} + \beta^{NC}_{ser}\theta^{NC}_{i} + u_{i,ser} \\
T_{i,sef} &= \alpha_{sef} + \beta^{NC}_{ses}\theta^{NC}_{i} + u_{i,sef} \\
T_{i,ses} &= \alpha_{ses} + \beta^{NC}_{ses}\theta^{NC}_{i} + u_{i,ses}
\end{aligned}
\tag{1}
$$

where:

- $T_{i,h}$ for $h \in \{math, cloze, ppvt, ser, sef, ses\}$: observed test scores and noncognitive skill measures.
- θ^{C}_{i} and θ^{NC}_{i}: unobserved/latent cognitive and noncognitive skill levels.
- $u_{i,h}$: error terms.

Model structure

- College completion decision:

$$D_i = \mathbb{1}[\alpha_D X_{iD} + \beta_D^C \theta_i^C + \beta_D^{NC} \theta_i^{NC} + u_{iD} > 0] \tag{2}$$

where:

  - $\mathbb{1}[A]$: an indicator function that equals one if A is true.

- Earnings for each channel of educational decision:

$$\begin{aligned} Y_{i,1} &= \alpha_{Y_1} X_{iY_1} + \beta_{Y_1}^C \theta_i^C + \beta_{Y_1}^{NC} \theta_i^{NC} + u_{iY_1} \text{ if } D_i = 1 \\ Y_{i,0} &= \alpha_{Y_0} X_{iY_0} + \beta_{Y_0}^C \theta_i^C + \beta_{Y_0}^{NC} \theta_i^{NC} + u_{iY_0} \text{ if } D_i = 0 \end{aligned} \tag{3}$$

where:

  - $Y_{i,0}$ and $Y_{i,1}$: hourly earnings measured at age 22 for D = 0, 1.
  - X_{iD} and X_{iY_D}: vectors of observed individual and household characteristics.

Estimation

- u_{ih} and u_{iY_D} are assumed to be normally distributed with mean zero and variances $\sigma^2_{u_h}$, $\sigma^2_{u_{Y_D}}$; u_{iD} is assumed to be logistically distributed.
- The latent factors are assumed to be distributed as a mixture of two normals:

$$f(\theta) = \sum_{c=1}^{2} \tau_c f(\theta|\mu_c, \Omega_c) \tag{4}$$

Where:

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

Estimation

- I estimate the model in one step by maximum likelihood using the minorization-maximization algorithm.
- The full model likelihood function:

$$L(\Psi) = \prod_{i=1}^{N} \int\int [f(T_i|\theta^C, \theta^{NC}) \times f(Y_{i,D=1}|X_{iY_{D=1}}, \theta^C, \theta^{NC})^D \times$$

$$f(Y_{i,D=0}|X_{iY_{D=0}}, \theta^C, \theta^{NC})^{1-D} \times Pr(D_i|X_D, \theta^C, \theta^{NC})] dF(\theta^C) dF(\theta^{NC}$$

$$= \prod_{i=1}^{N} \int f(T_i, D_i, Y_i|X_{i_D}, X_{iY_D}, \theta) f(\theta) d\theta \tag{5}$$

- The latent factors are assumed to be distributed as a mixture of two normals:

$$f(\theta) = \sum_{c=1}^{2} \tau_c f(\theta|\mu_c, \Omega_c) \tag{6}$$

Estimation

- Given the unobservable nature of the factors, the likelihood function is integrated over the distributions of these unobservable factors.
- I estimate the model in one step by maximum likelihood using the minorization-maximization algorithm.

Main results: Measurement system

	PPVT	Math	Cloze	Self-Esteem	Self-Efficacy	Self Respect
Panel A: Estimated parameters						
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)
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; *** 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.

Results: Probability of College Completion

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	-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)
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.

Results: 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)
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.

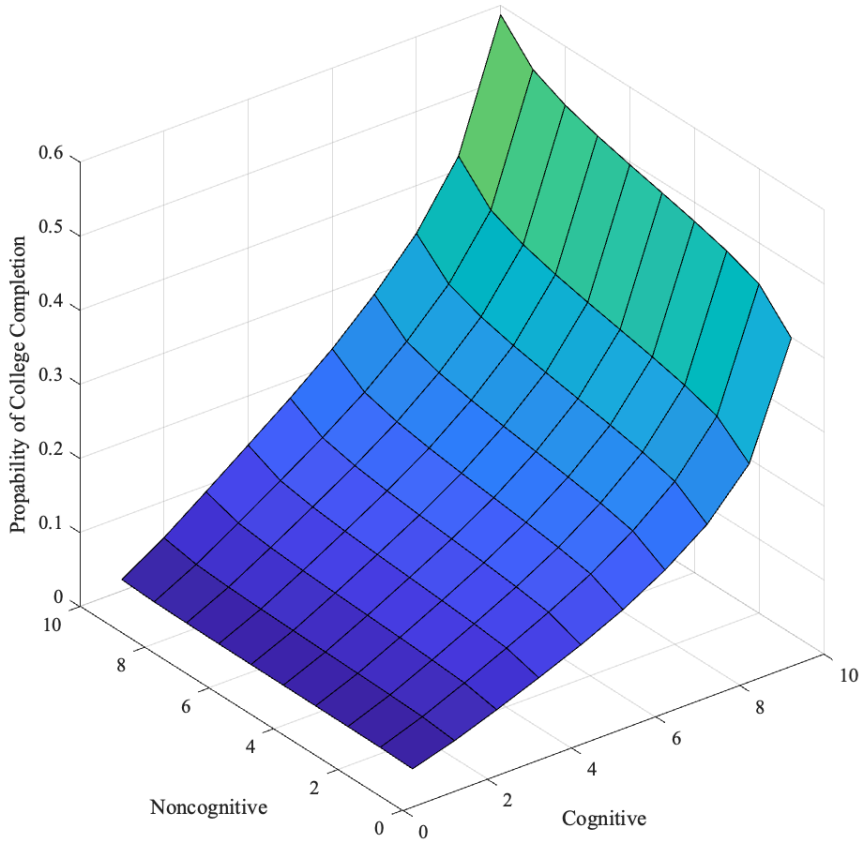
[a] Average Marginal Effects of Factors.

Treatment Effects

	Estimates
Average Treatment Effect $= E[Y_1\|X,\theta) - (Y_0\|X,\theta)]$	3.082^{*} (17.9%) (2.200)
Treatment Effect on the Treated $= E[Y_1\|X,\theta,D=1) - (Y_0\|X,\theta,D=1)]$	6.536^{***} (38%) (0.721)
Treatment Effect on the Untreated $= E[Y_0\|X,\theta,D=0) - (Y_1\|X,\theta,D=0)]$	-1.909 (-11.1%) (2.893)

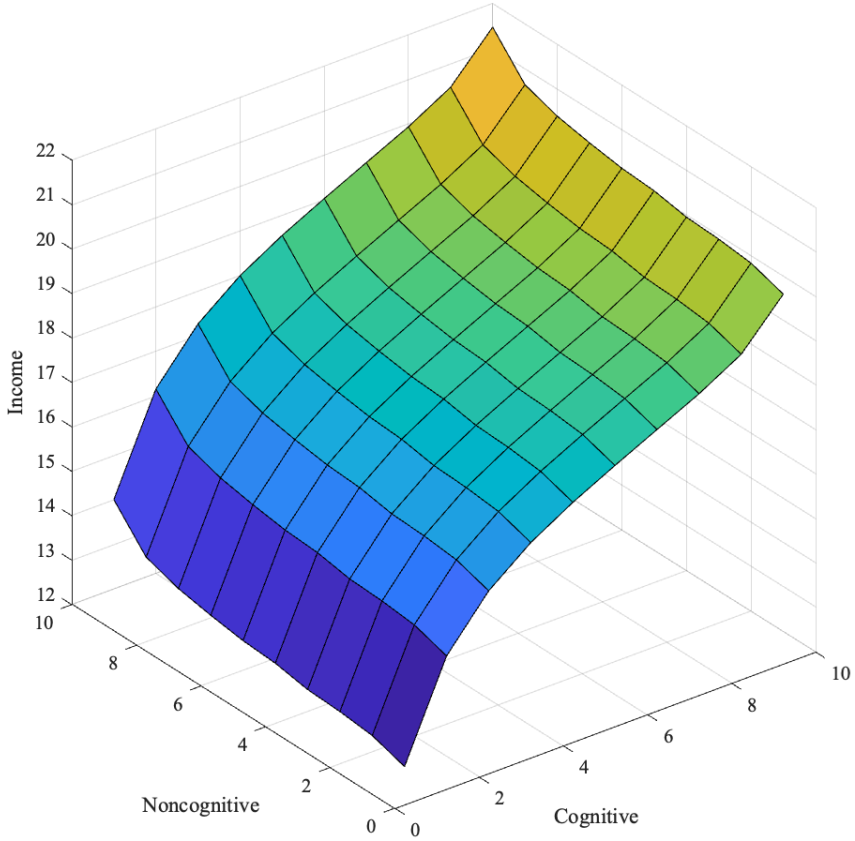
Notes: Standard errors in parentheses based on 100 bootstrap replications of the entire estimation process; *** p<0.01, ** p<0.05, * p<0.1.

Effects of Skills on Probability of College Completion



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.

Effects of Skills on Earnings



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.

Conclusions

- The effects of skills on earnings operate not only indirectly through the educational channel but also directly in the labor market.
- Among college graduates, noncognitive skills not only directly influence earnings in the labor market but also have indirect effects through educational choices.
- Among non-college graduates, cognitive skills are associated with better earnings.
- The effects of cognitive skills on earnings are stronger, mainly because of their stronger effects on educational choice.

Thank you for your attention!