The Long Shadow

Childhood Poverty and the Returns to Education

Ade Febriady[1,2,*] Agnieszka Postepska[1] Viola Angelini[1]

Asian Economic Development Conference 2026

13–14 June 2026 ▪ Hong Kong, China

[1]University of Groningen [2]World Bank (Consultant) [*]Presenter & corresponding author

a.febriady@rug.nl | a.febriady@worldbank.org | Available as GLO Discussion Paper 1731

Views are the author's own and do not represent the World Bank.

1 in 6 children live in extreme poverty —

half of the world's poorest are children.

Extreme poverty has a child's face.

World Bank & UNICEF (2023)

Why this matters

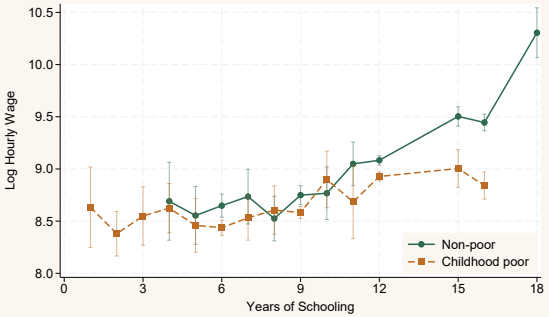
Education is meant to be **the great equalizer** — yet the return to a year of school is not:

1.5% **childhood poor** vs **6.8%** **everyone else**

return per additional year of schooling

- Expanding **access** is not enough
- The **"paradox of progress"** (Bourguignon 2005)
- The poor perceive **lower returns** (Boneva & Rauh 2017) — perhaps rightly
- **Mixed findings in SES gradients of the return** so far — mostly descriptive:
  Papanicolaou & Psacharopoulos 1979; Armitage & Sabot 1986; Cohn & Kiker 1986; Bartik & Hershbein 2018; Cheng et al. 2021

The raw fact, in the Indonesian setting



Mean log hourly wage by years of schooling. IFLS, ages 15–35.

The pattern

Wages rise for *everyone* — but **flatter** for the childhood poor; the gap *widens*.

Setting & data — Indonesia (IFLS)

- Access largely solved ⇒ we study the *return*, not attainment
- 5 waves (1993–2014); $N = 5{,}939$ wage workers aged 15–35 in 2014 (born 1979–99)
- Childhood poor = ever below \$2.15/day before age 15 — **54%**

Are the returns causal?

OLS bundles a year's true **payoff** with **who selects** into school.

Separating *return* from *selection* is the whole challenge.

Two sides of a coin: IV (2SLS) $\equiv$ Control Function (CF)

$$\text{wage:} \quad w_i = \beta x_i + \gamma\, \text{educ}_i + u_i$$

$$\text{schooling:} \quad \text{educ}_i = \delta x_i + \pi\, z_i + v_i$$

unobserved ability raises both $\Rightarrow \mathrm{Cov}(u_i, v_i) \neq 0 \Rightarrow$ OLS biased

IV / 2SLS

$$w_i = \beta x_i + \gamma\, \widehat{\text{educ}}_i + r_i$$

Control function

$$w_i = \beta x_i + \gamma\, \text{educ}_i + \boxed{\lambda\, \hat{v}_i} + e_i$$

same $\hat{\gamma}$ — numerically identical, Wooldridge 2015, traces to at least Hausman 1978

One catch: both routes **need an instrument** z — and in this context, we have none

e.g. school construction *(Duflo 2001)* predates our cohort (born 1979–99); the 1994 compulsory-schooling reform is weak *(Lewis 2020)*; parental background likely fails the exclusion restriction *(Card 1999)*.

No instrument? Klein & Vella (2010, *Econometrica*): identification via conditional heteroskedasticity

Recall the control function adds one term, λv_i (slide 6) — without an instrument it is collinear with schooling, so γ and λ are inseparable. **Klein & Vella's fix:** λ is not one fixed number — it **decomposes**: $\lambda = \rho \dfrac{S_u(x_i)}{S_v(x_i)}$

ρ: correlation of the unobservables (CCC) | S_u/S_v: relative spread, *varies* across people

Final estimating equation

$$w_i = \beta x_i + \gamma\,\text{educ}_i + \rho \underbrace{\frac{S_u(x_i)}{S_v(x_i)} v_i}_{\text{the source of identification}} + e_i.$$

S_u/S_v **varies** $\Rightarrow$ no longer collinear with schooling $\Rightarrow$ **identified, without an instrument**.

Multiplicative error structure + CCC ($E[u^* v^* \mid x] = \rho$, implicit in any IV). $\Rightarrow$ full derivation

We estimate v, S_v, S_u — what we hold fixed (covariates x) and use to predict the variance (predictors $z \subseteq x$)

Equation	Variables
Conditional means — covariates (x)	
Wage (w)	age, age^2, gender, ethnicity, birth region
Schooling (educ) $\rightarrow v$	age, age^2, gender, ethnicity, birth region
Conditional variances (S_j, $j = u, v$) — predictors ($z \subseteq x$)	
Schooling (S_v)	age, age^2, gender, ethnicity, birth region
Wage (S_u)	age, age^2, gender, ethnicity (birth region excluded)

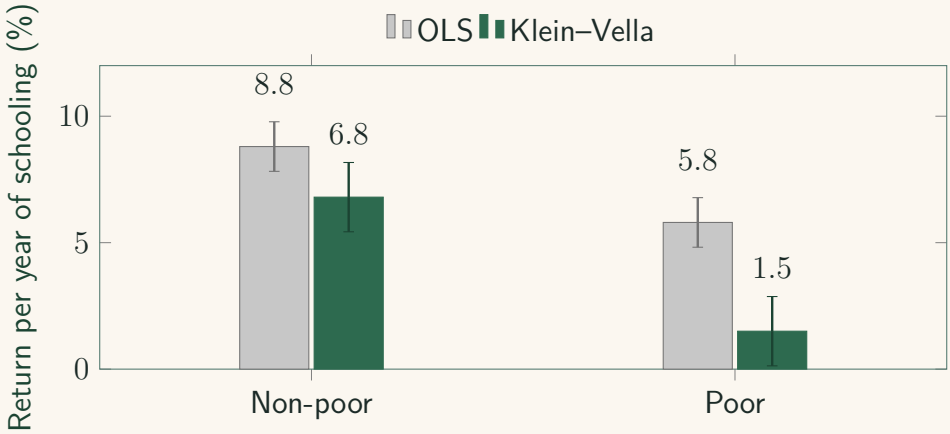
Age, not potential experience (a deterministic function of schooling). **Occupation** and **post-schooling cognitive skills** are *excluded* — outcomes of schooling (bad controls); used later, descriptively, for mechanisms. Here the variance predictors z are the same covariates x, only dropping birth region from S_u ($z \subseteq x$) — the heteroskedasticity that drives identification.

The childhood poor earn just 1.5% per year of schooling

	OLS			Klein–Vella		
	Non-poor	Poor	Diff	Non-poor	Poor	Diff
Years of schooling	0.088^{***}	0.058^{***}	0.029^{***}	**0.068^{***}**	**0.015^{**}**	**0.053^{***}**
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.010)
Selection $(S_u/S_v)\,\hat{v}$				0.077^{**}	**0.230^{***}**	-0.153^{***}
				(0.034)	(0.036)	(0.049)
N	2,708	3,231		2,708	3,231	

Bootstrapped SEs (500 reps). Diff = Non-poor − Poor. Controls: age, age^2, gender, ethnicity, birth region.

The correction collapses the poor's return



Bars: 95% CI. Each KV estimate is causal *within* its group; the gap is a comparison. Non-poor: $8.8 \to 6.8$; poor: $5.8 \to \mathbf{1.5}$.

The gap widens because selection is three times stronger among the poor

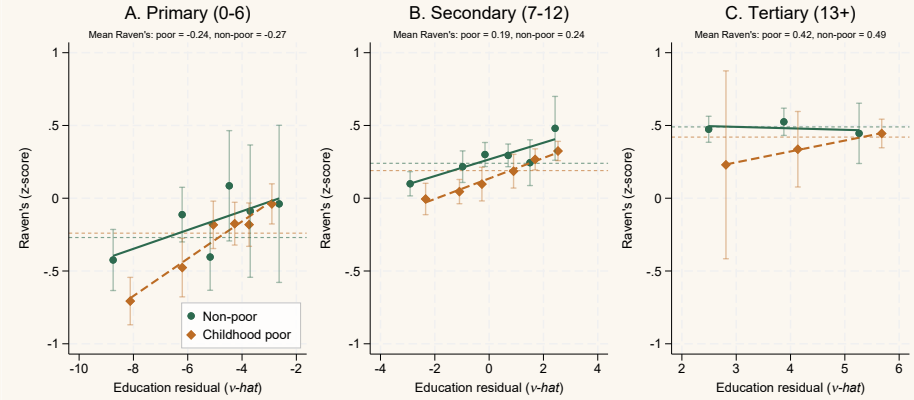
	Selection term $\hat{\rho}$
Non-poor	0.077^{**}
Poor	$\mathbf{0.230}^{***}$

$$\hat{\rho}_{\text{poor}} \approx 3 \times \hat{\rho}_{\text{non-poor}}$$

The intuition:

- Among the poor, only those with *exceptionally* favorable unobserved traits reach higher education
- OLS credits education with what is really the payoff to those traits
- So OLS **overstates** the return for the poor — and **understates** the true gap

Those who climb highest despite poverty are the most able — as selection predicts

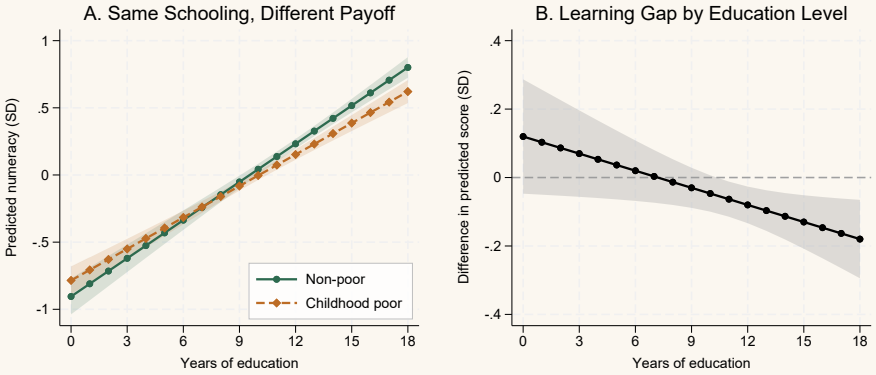


Mean Raven's z-score vs. bins of the education residual $\hat{v}$, by poverty status and education tier.

Why are returns lower for the poor?

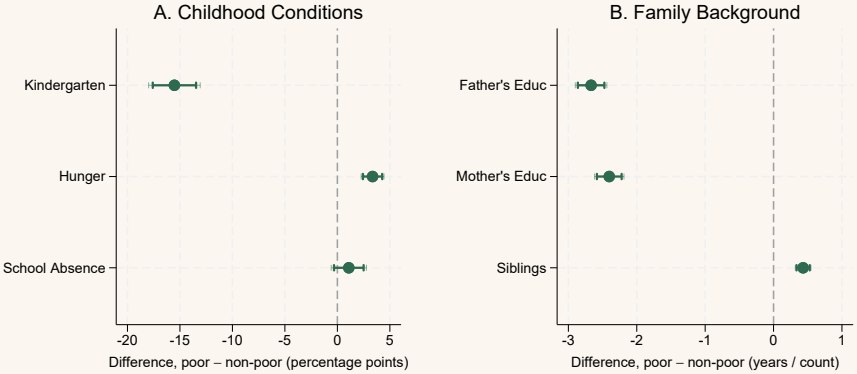
- Lower **skill acquisition** (at the same years of schooling)
- Lower chance of a **high-pay job** (at the same education and ability)

Mechanism I: the poor gain fewer skills from the same years of schooling



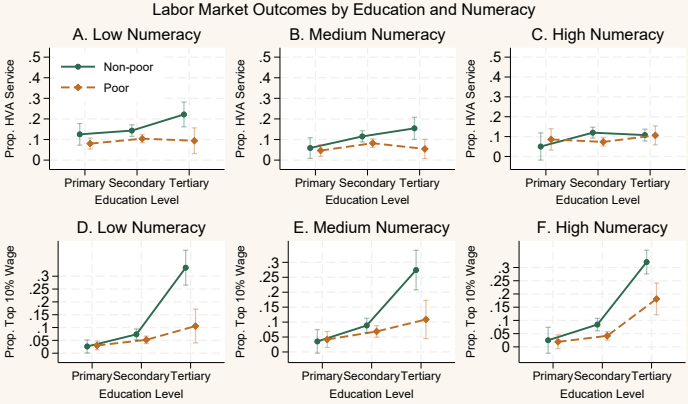
Predicted numeracy z-scores, education × poverty interaction. Panel B: poor − non-poor difference. Descriptive.

Why might they learn less? The poor start with weaker early-life foundations



Poor — non-poor differences from bivariate regressions. Panel A: childhood conditions; Panel B: family background. Descriptive.

Mechanism II: fewer high-paying jobs, even at equal numeracy



Row 1: high value-added service employment. Row 2: top wage decile. By education and numeracy tercile. Descriptive.

The gap survives every robustness check

Specification	KV gap (pp)
Baseline	**5.3*****
Alternative poverty lines ($0.8\times$, $1.2\times$ PL)	3.8–5.7***
Alternative wage-outlier treatments	4.2–6.1***
Birth region in wage variance function	8.1***
Restrict to ages 25–35	5.5***
Nonlinear returns (educ^2)	widens at higher educ
Continuous childhood PCE (second CF)	positive gradient***

Every specification confirms: the childhood poor earn substantially lower returns.

What it means

What this paper adds

- **Heterogeneous returns by childhood poverty** — first to address selection on *unobservables* (ability), not just observables, in the broader SES-gap returns literature [summary].[1]
- **A new mobility margin** — poverty lowers the *return* per year, not just *attainment* (Bellani & Bia 2019).
- **The "paradox of progress"** (Bourguignon 2005) — beyond convex returns, unequal access, and the tertiary-spending tilt (Gruber & Kosack 2014): lower returns at the *same* attainment.
- **SES enrolment gaps** — the poor perceive lower returns (Boneva & Rauh 2017); realized returns really are lower.

[1] Cheng et al. (2021) match on *observables* (US college returns) and find the opposite sign — low-SES earn the largest returns; matching cannot absorb the selection on unobservables ($\hat{\rho}$) we recover.

A few caveats

- **"Only wage workers, early-career."** Ability is near-identical for the self-employed; ages 25–35 barely move the gap.
- **"KV is only partly testable."** True — but even OLS shows a significant gap, a likely *lower bound*; robust across specs.
- **"The mechanisms aren't causal."** Correct — descriptive correlates, not identified channels.

Childhood poverty casts a long shadow on the value of education

1.5% per year against **6.8%** — and the correction *widens* the gap.

What this changes

- Access alone won't equalize — unequal returns can *widen* gaps
- The lever isn't *reaching* school, but **what they gain there**

Education can still be the equalizer — if we close the gap in what a year delivers.

For those who grew up poor,

an extra year of school pays a **fraction**

of what it pays everyone else.

Thank you! 多謝!

Ade Febriady a.febriady@rug.nl

Available as GLO Discussion Paper 1731 | LinkedIn

Backup Slides

Returns to schooling by SES: mixed findings, mostly OLS

Study	SES measure	Method	Returns by SES
Papanicolaou & Psacharopoulos (1979) UK	Father's occupation	OLS	**Decline** with SES
Armitage & Sabot (1986) Kenya/Tanzania	Parental education	OLS	**Increase** with SES
Cohn & Kiker (1986) US	Father's occupation (+IQ)	OLS	**Similar** across SES
Bartik & Hershbein (2018) US college	Family income (13–17)	OLS	**Increase** with SES
Cheng et al. (2021) US college	Parental income, educ, ability	Matching	**Decline** with SES
This paper Indonesia	Childhood poverty (<$2.15/day)	**CF (KV)**	**Decline** — unobservables

Mostly OLS or observable-based matching; we add identification that addresses selection on *unobservables* (ability).

↩ back

Klein–Vella in full: error structure and derivation

Step 1. Impose a **multiplicative structure** on the structural errors:

$$u_i = S_u(x_i)\, u_i^*, \quad v_i = S_v(x_i)\, v_i^*, \qquad \overbrace{E[u_i^* v_i^* \mid x_i] = \rho}^{\text{constant conditional correlation (CCC)}} \quad .$$

Step 2. The endogeneity parameter then follows directly:

$$\lambda \;=\; \frac{\mathrm{Cov}(u,v)}{\mathrm{Var}(v)} \;\xrightarrow{\text{CCC}}\; \rho\,\frac{\sigma_u(x_i)}{\sigma_v(x_i)} \;\xrightarrow{\sigma_j = S_j}\; \rho\,\frac{S_u(x_i)}{S_v(x_i)}.$$

Step 3. Substitute into the wage equation:

$$w_i = \beta x_i + \gamma\,\mathsf{educ}_i + \rho \underbrace{\frac{S_u(x_i)}{S_v(x_i)}\, v_i}_{\text{the source of identification}} + e_i.$$

The scaling $S_u(x)/S_v(x)$ varies across observable groups — no excluded instrument is required. ↩ back

Klein–Vella in full: estimation and assumptions

Two-step parametric estimation (following Farr 2013):

1. OLS of the education equation $\Rightarrow$ residual $\hat{v}_i$.
2. Parametric variance functions $S_{ji}^2 = \exp(z_{ji}\theta_j)$, $j = u, v$; recover $\widehat{S}_u, \widehat{S}_v$. Predictors z need not equal x.
3. Insert $(\widehat{S}_u / \widehat{S}_v)\,\hat{v}_i$ into the wage equation; estimate γ and ρ. Bootstrap SEs, 500 reps.

Identifying assumptions:

- **Constant correlation (CCC):** $E[u_i^* v_i^* \mid x_i] = \rho$. Untestable, but implicit in standard 2SLS; we allow ρ to differ *across* poverty groups.
- **Heteroskedasticity:** S_u / S_v not constant. Tested — Breusch–Pagan and White reject homoskedasticity.

Full estimation results

	OLS				Klein-Vella			
	All	Non-poor	Poor	Diff	All	Non-poor	Poor	Diff
Years of schooling	0.080***	0.088***	0.058***	0.029***	0.042***	0.068***	0.015**	0.053***
	(0.003)	(0.005)	(0.005)	(0.007)	(0.008)	(0.007)	(0.007)	(0.010)
$(S_u/S_v) \cdot \hat{v}$					0.183***	0.077**	0.230***	-0.153***
					(0.037)	(0.034)	(0.036)	(0.049)
Observations	5,939	2,708	3,231		5,939	2,708	3,231	
Heteroskedasticity Tests (p-values):								
BP test (wage)					0.000	0.000	0.000	
BP test (educ)					0.000	0.000	0.000	
White test (wage)					0.000	0.024	0.000	
White test (educ)					0.000	0.000	0.000	

Bootstrapped SEs (500 reps). Controls: age, age^2, gender, ethnicity, birth region.

First stage: education equation

	All	Non-poor	Poor
Constant	-3.315***	-6.266***	-3.655***
	(0.492)	(0.525)	(0.528)
Female	0.978***	1.039***	0.725***
	(0.079)	(0.081)	(0.081)
Ethnicity (Javanese)	0.006	0.084	-0.136
	(0.082)	(0.083)	(0.084)
Age	1.076***	1.293***	1.078***
	(0.032)	(0.033)	(0.034)
Age squared	-0.020***	-0.023***	-0.021***
	(0.000)	(0.001)	(0.001)
Java - Village	-1.025***	-0.493***	-0.280**
	(0.129)	(0.131)	(0.136)
Java - Small town	-0.458***	0.097	-0.023
	(0.163)	(0.152)	(0.166)
Java - Big city	0.197	0.737***	0.295
	(0.221)	(0.235)	(0.229)
Off-Java - Village	-0.799***	-0.185	0.012
	(0.131)	(0.135)	(0.138)
Off-Java - Small town	0.148	0.804***	0.307*
	(0.162)	(0.151)	(0.160)
Off-Java - Big city	0.670***	0.708***	1.046***
	(0.203)	(0.208)	(0.206)
Observations	5,939	2,708	3,231
R^2	0.064	0.089	0.041
F-statistic	40.27	26.21	13.77
White test p	0.000	0.000	0.000
Breusch-Pagan p	0.000	0.000	0.000

Nonlinear returns: the gap widens at higher attainment

	Log hourly wage		
	Non-poor	Poor	Diff (ME only)
Panel A: Coefficient Estimates			
Years of schooling	-0.025**	0.016	-0.041***
	(0.012)	(0.011)	(0.016)
Years of schooling^2	0.004***	-0.000	0.004***
	(0.001)	(0.001)	(0.001)
$(S_u/S_v) \cdot \hat{v}$	0.086**	0.227***	
	(0.036)	(0.036)	
Panel B: Marginal Effects at Specified Education Levels			
Observations	2,708	3,231	
Joint p-value (educ, educ^2)	0.000	0.356	
ME at 6 years (SE)	0.025 (0.013)	0.016 (0.013)	0.010 (0.018)
ME at 9 years (SE)	0.051 (0.015)	0.015 (0.014)	0.035* (0.021)
ME at 12 years (SE)	0.076 (0.017)	0.015 (0.017)	0.061** (0.024)
ME at 16 years (SE)	0.110 (0.021)	0.015 (0.020)	0.095*** (0.029)

No differential ability-based selection into wage work

	Non-poor			Poor		
	Wage worker	Self-empl.	Norm. diff.	Wage worker	Self-empl.	Norm. diff.
Years of schooling	11.77 (3.39)	10.58 (3.43)	0.35***	9.83 (3.46)	8.70 (3.40)	0.33***
Raven's z-score	0.21 (0.94)	0.17 (0.89)	0.04	0.08 (0.92)	0.05 (0.87)	0.04
Numeracy z-score	0.18 (1.06)	0.15 (1.01)	0.04	-0.04 (0.95)	-0.07 (0.91)	0.04
Observations	3,011	1,144		3,605	1,594	
N (cognitive)	3,009	1,143		3,604	1,593	

Norm. diff. $= (\bar{X}_W - \bar{X}_S)/\sqrt{(s_W^2 + s_S^2)/2}$. Cognitive ability nearly identical across employment types.

Robustness: alternative poverty-line definitions

	Main ($2.15)			0.8×PL ($1.72)			1.2×PL ($2.58)		
	Non-poor	Poor	Diff	Non-poor	Poor	Diff	Non-poor	Poor	Diff
Years of schooling	0.068***	0.015**	0.053***	0.050***	0.012	0.038***	0.087***	0.031***	0.057***
	(0.007)	(0.007)	(0.010)	(0.007)	(0.008)	(0.010)	(0.007)	(0.007)	(0.010)
$(S_u/S_v) \cdot \hat{v}$	0.077**	0.230***	-0.153***	0.132***	0.231***	-0.099*	0.017	0.173***	-0.156***
	(0.034)	(0.036)	(0.049)	(0.034)	(0.038)	(0.051)	(0.036)	(0.036)	(0.051)
Observations	2,708	3,231		3,537	2,402		2,040	3,899	