Sibs, schools or sorting? Accounting for learning inequalities across East Africa

Paul Anand Jere Behrman Hai-Anh Dang Sam Jones

Open University // UPenn // World Bank // UNU-WIDER

19 March 2019

Outline

(1) Motivation

Low average levels of achievement

Widespread concerns about the ‘global learning crisis’ (UNESCO, 2013).

> “... many students are learning close to nothing in school”

(Study Group on Measuring Learning Outcomes, 2013: 1).

$\Longrightarrow$ Dominant discussion around *average* outcomes.

Our starting point: **learning inequalities also important** (e.g., social justice perspective).

Multiple possible sources of inequalities:

- Household factors (e.g., wealth, nutrition; Walker et al., 2011)
- School/teacher quality (Bietenbeck et al., 2017; Bold et al., 2018)
- Sorting (clustering) (Hanushek & Yilmaz, 2007)

... Previous studies have looked at these, but typically separately and/or only using observed proxies.

(2) This paper

What does this paper do?

Aim: quantify the contribution of different sources of inequalities in *opportunity to learn*.

We contribute:

- Factor-based decomposition, covering multiple *latent* factors of interest
- New estimation approach, permits alternative assumptions about factor covariance (sorting) in a single framework
- Unique coverage: >1 million children in 3 East African countries; permits comparison of differences across space

Limitations:

- Strictly speaking, bounds on variance contributions
- Imperfect identification of schools
- Floor + ceiling effects in test scores

= Diagnostic & exploratory

(3) Framework

Framework

Starting point is an educational production function.

A generic factor model:

$$t_{ijk} = f(e_i, h_j, s_k) \tag{1}$$

e_i: individual & idiosyncratic effects

h_j: household or sibling effects

s_k: schooling effects (external to household)

To proceed, we need:

1. A measure of inequality linked to (1)
2. Place structure on f (esp., how do h and s relate?)

Measure of inequality

Measure inequality via the **variance** of t.

Widely used in existing literature.

Attractive properties:

1. Ordinally invariant to test score standardization (Ferreira & Gignoux, 2014)
2. Factor decomposability (Shorrocks, 1982)
3. Sub-group decomposability (Chakravarty, 2001)

Framework

How to think about the relation between different effects?
... assuming individual effects are independent (for now).

Model		Score level	Score variance
1	Restricted linear	$h_j + s_k + e_{ijk}$	$\sigma_h^2 + \sigma_s^2 + \sigma_e^2$
2	Unrestricted linear	$h_j + s_k + e_{ijk}$	$\sigma_h^2 + \sigma_s^2 + 2\Sigma_{hs} + \sigma_e^2$
3	Household upper-bound	$(1+\gamma)h_j + \nu_k + e_{ijk}$	$(1+\gamma)^2\sigma_h^2 + \sigma_\nu^2 + \sigma_e^2$
4	School upper-bound	$(1+\theta)s_k + \omega_j + e_{ijk}$	$(1+\theta)^2\sigma_s^2 + \sigma_\omega^2 + \sigma_e^2$

Variance decomposition (full)

$$t_{ijk} = a_i + h_j + s_k + \epsilon_{ijk}, \qquad \text{(2a)}$$
$$\text{where } \; a_i = x_i'\beta$$

$$\mathrm{Var}(t_{ijk}) \equiv \sigma_t^2 = \sigma_a^2 + \sigma_h^2 + \sigma_s^2 + 2\Sigma_{hs} + 2\Sigma_{ah} + 2\Sigma_{as} + \sigma_\epsilon^2 \qquad \text{(2b)}$$

$$\implies 1 = \frac{\sigma_a^2}{\sigma_t^2} + \frac{\sigma_h^2}{\sigma_t^2} + \frac{\sigma_s^2}{\sigma_t^2} + \frac{2\Sigma_{hs}}{\sigma_t^2} + \frac{2\Sigma_{ah}}{\sigma_t^2} + \frac{2\Sigma_{as}}{\sigma_t^2} + \frac{\sigma_\epsilon^2}{\sigma_t^2} \qquad \text{(2c)}$$

(4) Methods

Decomposition methods

Different estimation methods apply to specific models:

- Restricted model : REML (`lmer`)
- Upper bound models : one-way fixed effect
- Unrestricted model : iterative FE algorithms (`reghdfe`)

Problems:

- Fixed effects estimated with noise := shrinkage required
- Mechanical negative correlation with multi-way effects (bias)

Dealing with 'limited mobility bias'

Issue raised in seminal paper by Abowd et al. (1999), which considers firm and worker effects. But arises in other settings.

Core problem is that both h_j and s_k are estimated by taking averages over residuals. ... So, estimates of the two effects depend on *which* effect is estimated first.

Example, assume (h, s) share a common 'mobility group' effect (here, by location), so true model is:

$$t_{ijk} = h_{j(v)} + s_{k(v)} + \mu_v + e_{ijk}$$

Näive estimate of (1st) household effect & (2nd) school effect:

$$\hat{h}_j = \mu_v + h_j + \bar{s}_k + \bar{e}_{ijk}$$

$$\implies \hat{s}_k = \frac{1}{N_k} \sum_{i|K=k} \left(t_{ijk} - \hat{h}_j \right) = \frac{1}{N_k} \sum_{i|K=k} (s_k - \bar{s}_k + e_{ijk} - \bar{e}_{ijk})$$

What to do?

No consensus. Work in progress.

A priori, we don't know where to allocate the common effect.

Our approach: make the initialization (normalization) assumption of latent effects transparent.

Choose $\pi \in [0, 1]$:

$$\hat{h}_{j(0)} = \frac{1}{N_j} \sum_{i|J=j} \left(t_{iJ} - \frac{\pi}{N_k} \sum_{i|K=k} t_{iK} \right) \tag{3a}$$

$$\hat{s}_{k(0)} = \frac{1}{N_k} \sum_{i|K=k} \left(t_{iK} - \frac{1-\pi}{N_j} \sum_{i|J=j} t_{iJ} \right) \tag{3b}$$

Midpoint, $\pi = 0.5 \implies$ agnostic initialization; equally partitions the shared effect (μ_v). Minimizes the mechanical bias.

Summary

- Standard variance decomposition
- Estimate multi-way fixed effects via iterative algorithm
- Use alternative initializations to capture different assumptions
- Single framework to estimate all effects of interest

(5) Data

Uwezo surveys

Large-scale household-based surveys of learning since 2010, following Pratham/ASER model (India, Pakistan).

Representative down to district level (370 in total).

All children in household aged 6-16 tested at Grade 2 level.

– Kenya & Tanzania: English, Swahili, Math

– Uganda: English, Math

Pool all five survey rounds (2011-2015) $\Longrightarrow$ > 1 million obs.

Outcome = synthetic overall test score, standardized by country, survey year & age. [$\mu_t = 0$; $\sigma_t = 1$]

Sample & effect definitions

Sample : children of primary school age 6-13 years.

Household fixed effect (h_j): residential unit.

School effect (s_k), four categories in each location (village):

1. Out of school children (excl. those already in secondary)
2. Children attending a specific public primary school
3. Children attending other public primary school(s)
4. Children attending private primary schools

Identification of separate effects from crossed units (sibs attending different schools).

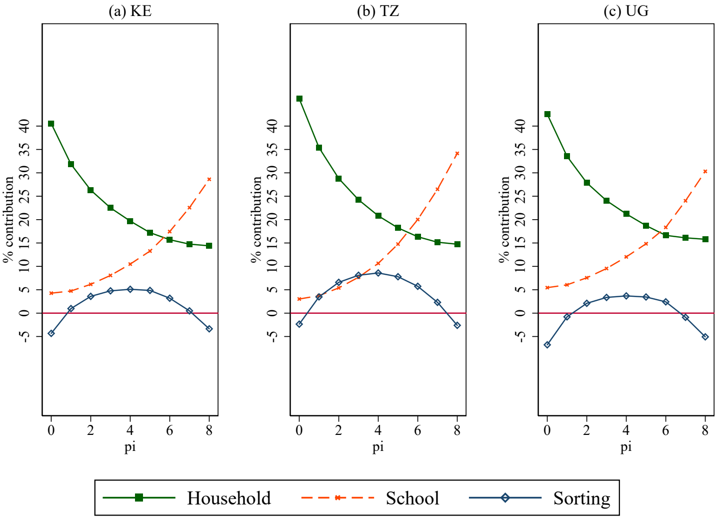
All singleton units removed.

Country & Region		Index count i	j	k	Schooling status Enrol.	Match	Private
KE	Central	25,363	11,067	4,761	96.7	44.3	25.9
	Coast	34,689	13,353	4,333	85.0	41.7	17.4
	Eastern	56,944	23,182	7,598	94.5	59.8	6.4
	North Eastern	40,281	14,500	3,038	79.7	55.2	3.4
	Nyanza	52,693	20,896	7,488	91.7	50.6	14.2
	Rift Valley	119,316	46,922	14,493	89.8	51.2	12.0
	Western	57,236	22,076	7,151	92.3	52.9	8.0
	All	386,522	151,996	48,862	91.1	50.8	13.2
TZ	Arusha	33,517	14,129	4,453	89.7	51.9	5.7
	Dar Es Salaam	14,101	5,994	2,078	92.9	45.7	6.6
	Iringa	30,082	13,178	4,357	87.5	61.3	3.2
	Kagera	33,146	13,286	4,239	83.1	52.8	3.6
	Kigoma	21,461	8,885	2,657	82.5	53.5	3.3
	Ruvuma	17,979	7,979	2,875	89.6	64.2	3.2
	Singida	20,487	8,796	2,665	87.1	61.6	2.6
	Tabora	33,738	13,276	3,888	79.0	51.2	3.1
	Tanga	25,075	10,538	3,225	87.6	59.9	3.4
	All	229,586	96,061	30,437	85.9	55.3	3.8
UG	Central	46,514	16,967	6,019	95.1	24.4	49.1
	Eastern	86,771	30,558	8,709	95.7	44.9	23.8
	Northern	72,333	26,661	6,804	87.2	53.5	6.3
	Western	52,774	20,074	6,746	94.0	36.9	28.3
	All	258,392	94,260	28,278	93.4	38.9	28.5

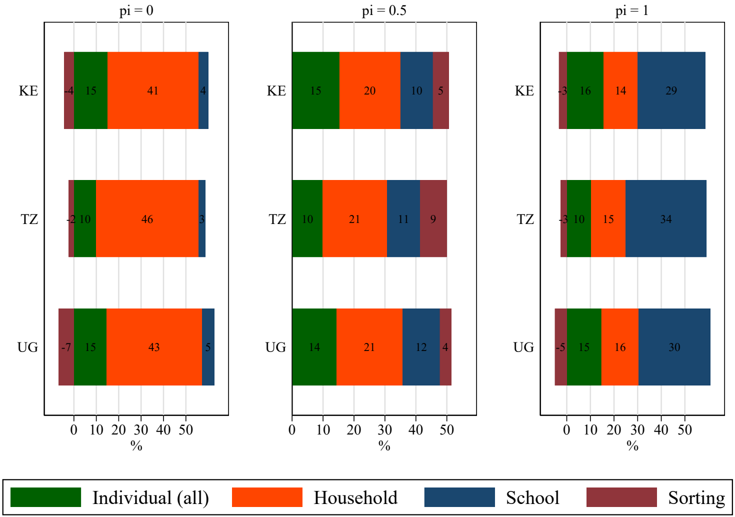
(6) Results

Estimator performs as expected

$\pi = 0$: household upper bound ; $\pi = 1$: school upper bound; $\pi = 0.5$: midpoint



Relative variance shares, by estimator
(Equivalent to absolute contributions, since $\sigma_t^2 = 1$)



Detailed results, absolute contributions, Kenya

Reported in sd (σ) units

Part	(a) $\pi = 0$	(b) $\pi = 0.5$	(c) $\pi = 1$	(d) UBH	(e) UBS	(f) MLM
$\sigma^2_{a^o}$	0.32	0.32	0.32	0.34	0.33	0.35
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ^2_h	0.64	0.44	0.38	0.61	0.33	0.32
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ^2_s	0.21	0.32	0.53	0.05	0.48	0.38
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$2\Sigma_{hs}$	-0.21	0.23	-0.18	0.07	0.15	0.24
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$2\Sigma_{a^o h}$	0.18	0.15	0.12	0.21	0.12	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$2\Sigma_{a^o s}$	0.13	0.16	0.19	0.05	0.20	0.18
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ^2_e	0.67	0.70	0.67	0.68	0.69	0.72
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ^2_t	1.00	1.00	1.00	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ρ_{hs}	-0.16	0.18	-0.08	0.07	0.07	0.24
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Detailed results, relative contributions, Kenya

Part	(a) $\pi = 0$	(b) $\pi = 0.5$	(c) $\pi = 1$	(d) UBH	(e) UBS	(f) MLM
$\sigma^2_{a^o}$	10.39	10.42	10.51	11.61	10.79	12.56
	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.11)
σ^2_h	40.55	19.69	14.40	36.69	10.58	10.22
	(0.27)	(0.19)	(0.16)	(0.26)	(0.14)	(0.14)
σ^2_s	4.27	10.49	28.62	0.26	23.35	14.53
	(0.14)	(0.22)	(0.36)	(0.03)	(0.33)	(0.26)
$2\Sigma_{hs}$	-4.33	5.10	-3.35	0.43	2.14	5.79
	(0.11)	(0.12)	(0.10)	(0.03)	(0.08)	(0.13)
$2\Sigma_{a^o h}$	3.17	2.20	1.40	4.22	1.33	1.65
	(0.07)	(0.05)	(0.04)	(0.08)	(0.04)	(0.05)
$2\Sigma_{a^o s}$	1.58	2.68	3.70	0.22	3.85	3.08
	(0.06)	(0.07)	(0.09)	(0.02)	(0.09)	(0.08)
σ^2_e	44.38	49.41	44.72	46.56	47.97	52.17
	(0.21)	(0.23)	(0.22)	(0.22)	(0.22)	(0.23)
σ^2_t	100.00	100.00	100.00	100.00	100.00	100.00
	.	.	.	.	.	.
ρ_{hs}	.	.	.	.	.	.
	.	.	.	.	.	.

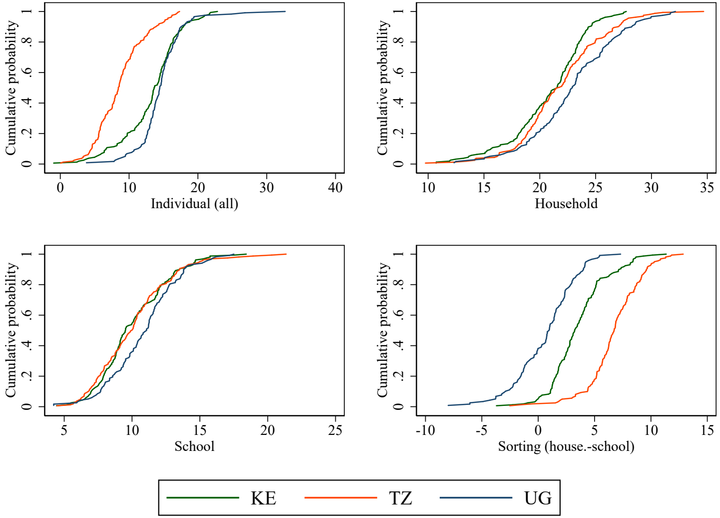
Sub-group decompositions, Kenya

Relative contributions (in %), $\pi = 0.5$

Strata		$\sigma^2_{a^o}+$	σ^2_h	σ^2_s	$2\Sigma_{hs}$	σ^2_e
Sch. cat.	0	2.87	21.91	17.22	5.37	52.63
	1	9.91	20.99	8.83	5.65	54.63
	2	10.73	21.63	9.96	2.51	55.17
	3	12.52	23.13	12.64	0.21	51.50
Private	0	10.21	21.22	9.23	4.53	54.80
	1	12.52	23.13	12.64	0.21	51.50
Female	0	9.92	19.62	10.32	5.17	54.97
	1	10.76	19.93	10.76	5.06	53.49
Age	6	5.38	20.53	11.99	4.43	57.66
	9	9.94	19.18	9.80	5.64	55.42
	12	11.21	18.34	9.54	4.81	56.10
SES tercile	1	9.08	18.89	10.83	5.12	56.08
	2	10.48	20.50	9.79	3.68	55.55
	3	13.18	21.14	11.07	3.89	50.73
Mother edu.	0	9.34	18.26	11.57	6.17	54.66
	1	11.49	21.08	9.89	3.96	53.58
All	.	10.54	19.69	10.49	5.10	54.17

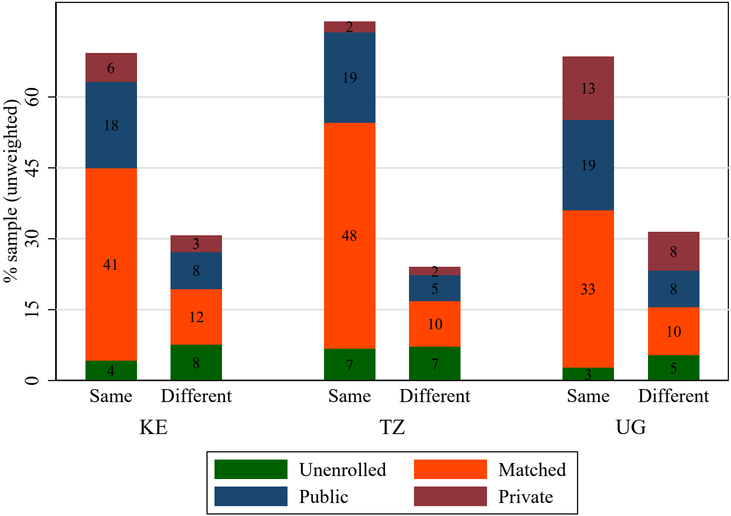
Spatial variation

Distributions of relative contributions by district, , $\pi = 0.5$

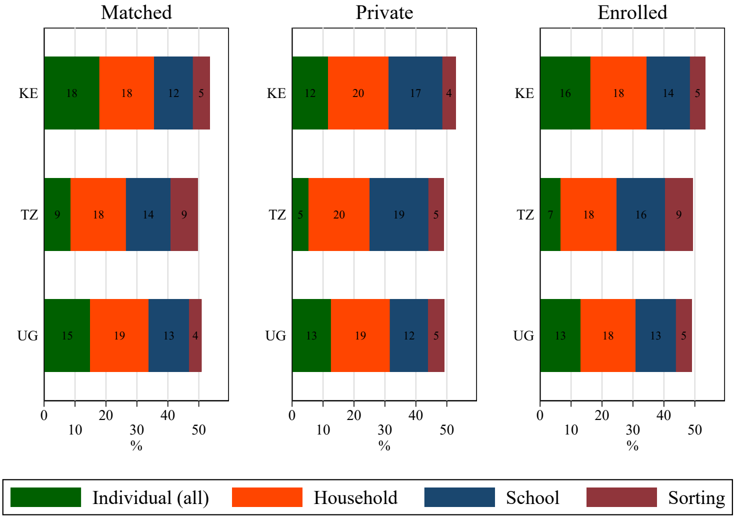


(7) Robustness

Identification: what % of the sample?

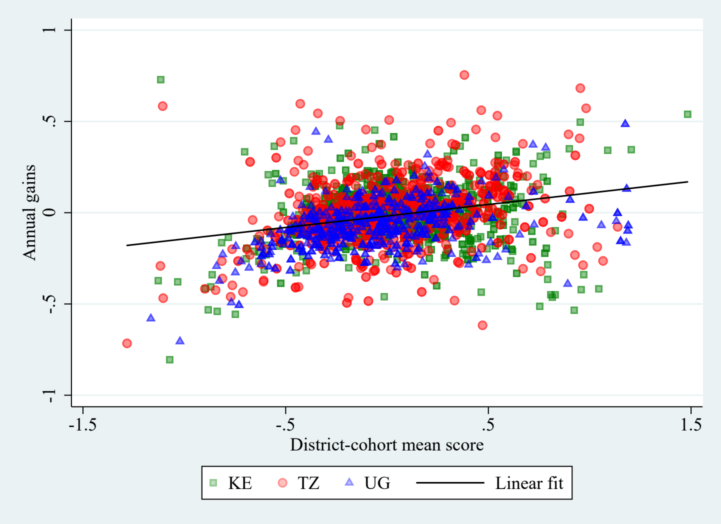


Robustness, only non-homogeneous households



District-level gains vs means (birth cohorts)

An alternative angle on sorting



Mean – slope correlation: TZ $>$ KE $>$ UG.

(8) Conclusion

Conclusions

1. We documented sources of learning inequalities in E. Africa.
2. We found IEO driven by various factors : *not* a zero contribution of schools *not* uniquely due to household effects
3. Variation between schools is material; and this is aggravated by positive material sorting, esp. in specific locations

$\Longrightarrow$ Move from 10th to 90th percentile of school distribution implies an expected test score difference of $> 0.66\sigma$

$\Longrightarrow$ Magnitudes much larger than teacher content knowledge effects, suggesting school management + other factors key

4. Schools do help equalise outcomes relative to no schooling
5. Agenda: policies to reduce variation in school quality *as well as* raise average quality