Teacher Effectiveness in Africa: Longitudinal and Causal Estimates

Julie Buhl-Wiggers (Copenhagen Business School)
Jason Kerwin (UMN)
Jeffrey Smith (Wisconsin)
Rebecca Thornton (UIUC)

NBER SI 2021 Education

July 28, 2021

How important are teachers in Africa?

- Teacher value-added (TVA) is a crucial driver of
  - Test scores (Hanushek and Rivkin 2010, Araujo et al. 2016)
  - Employment and wages (Koedel et al. 2015)

- Potentially different in African schools
  - Limited resources
  - Hiring/training differences

- Teachers could matter more or less than in developed world
  - Skills needed to deal with conditions $\implies$ more important
  - Constraints make success impossible $\implies$ less important

- No previous teacher value-added estimates for Africa

African classrooms are challenging teaching environments

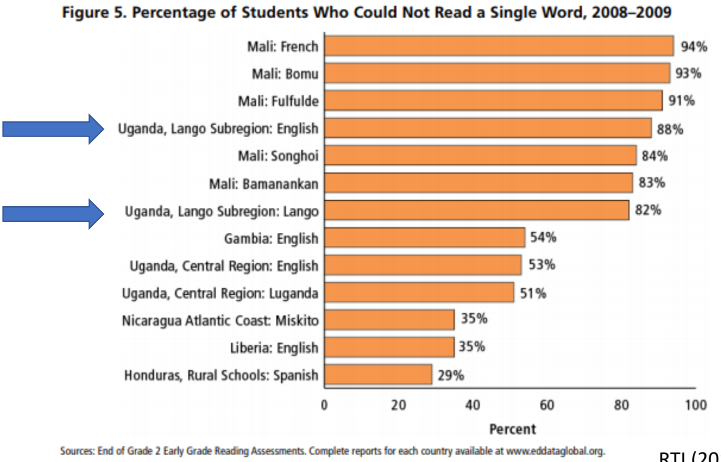
A typical classroom in Uganda
(and in sub-Saharan Africa)

- Schools in Uganda have very limited resources
  - Mean class size: 109
  - Average attendance rate: 45%
  - Spending per child per year: $55

Quality of teaching is generally poor

- Teaching practices are antiquated in Uganda
  - Old school call-and-response pedagogy (Ssentanda 2014)
  - Instruction in English rather than the local language (Ssentanda et al. 2016)
- Poor teacher training, limited teacher effort across seven African countries (Bold et al. 2017)
  - Less than 40% of teachers meet minimum knowledge standards for language education or general pedagogy
    - Uganda's teacher training is not very applicable to the classroom (Hardman et al. 2011)
  - Teachers actually teach for just 50 percent of scheduled class time, < 3 hrs/day

And students don't learn much



RTI (2011)

Will improving teacher quality help address these problems?

- Mounting evidence suggests teacher quality is a key determinant of learning
- Two approaches:
  1. **Teacher Effectiveness:** Estimate teacher value-added (TVA) and find that variation in TVA explains a substantial part of the variation in test scores. (e.g. Chetty et al. 2014, Araujo et al. 2016, Azam & Kingdon 2015, Bau & Das 2020)
  2. **Program Evaluation:** Interventions involving teacher training are some of the most effective. (Kremer et al. 2013, Glewwe & Muralidharan 2015, Ganimian & Murnane 2014, McEwan 2015, Evans & Popova 2016)
- This is the first paper to integrate these two approaches

This Paper

1. How effective are Ugandan teachers? Estimate TVA
   - We provide the first estimates of TVA in Africa

2. Which teachers are effective? Correlate teachers' effectiveness with their characteristics

3. How does a teacher training-based intervention affect teacher quality? Measure the impact of a randomized intervention on TVA

Preview of Results

- A 1 SD increase in teacher effectiveness increases student learning by at least 0.18 SDs in local-language reading and 0.20 SDs in English reading—a bigger effect than in the US

Preview of Results

- A 1 SD increase in teacher effectiveness increases student learning by at least 0.18 SDs in local-language reading and 0.20 SDs in English reading—a bigger effect than in the US
- Moving from a 10^{th} percentile teacher to a 90^{th} percentile teacher raises test scores by as much as the most-effective education interventions

Preview of Results

- A 1 SD increase in teacher effectiveness increases student learning by at least 0.18 SDs in local-language reading and 0.20 SDs in English reading—a bigger effect than in the US
- Moving from a 10th percentile teacher to a 90th percentile teacher raises test scores by as much as the most-effective education interventions
- Teacher effectiveness is essentially uncorrelated with observed exogenous characteristics

Preview of Results

- A 1 SD increase in teacher effectiveness increases student learning by at least 0.18 SDs in local-language reading and 0.20 SDs in English reading—a bigger effect than in the US
- Moving from a 10th percentile teacher to a 90th percentile teacher raises test scores by as much as the most-effective education interventions
- Teacher effectiveness is essentially uncorrelated with observed exogenous characteristics
- A teacher training-focused intervention increases the spread of the TVA distribution (likely by making the good teachers better)

The Northern Uganda Literacy Project (NULP)

- Program developed by Mango Tree, a Ugandan education firm
- Two versions: full-cost and reduced-cost

The Northern Uganda Literacy Project (NULP)

- Program developed by Mango Tree, a Ugandan education firm
- Two versions: full-cost and reduced-cost
- Full-cost: local language-first instruction, detailed lesson plans / scripts, training and monitoring by Mango Tree staff, primers, readers and slates for all students, wall clocks. Intervention runs from P1 to P3.

The Northern Uganda Literacy Project (NULP)

- Program developed by Mango Tree, a Ugandan education firm
- Two versions: full-cost and reduced-cost
- Full-cost: local language-first instruction, detailed lesson plans / scripts, training and monitoring by Mango Tree staff, primers, readers and slates for all students, wall clocks. Intervention runs from P1 to P3.
- Reduced-cost: Same as full-cost but "cascade" (training-of-trainers) training and monitoring by government staff, slates and clocks only for a subset of schools
    - Designed to represent how program could be scaled up

The Northern Uganda Literacy Project (NULP)

- Program developed by Mango Tree, a Ugandan education firm
- Two versions: full-cost and reduced-cost
- Full-cost: local language-first instruction, detailed lesson plans / scripts, training and monitoring by Mango Tree staff, primers, readers and slates for all students, wall clocks. Intervention runs from P1 to P3.
- Reduced-cost: Same as full-cost but "cascade" (training-of-trainers) training and monitoring by government staff, slates and clocks only for a subset of schools
    - Designed to represent how program could be scaled up
- In separate papers, we study:
    - the program's average effects at the end of P1 (Kerwin and Thornton 2021)
    - end-of-P3 effects and the scale-up of the program (Buhl-Wiggers et al., in progress)
    - heterogeneity in the treatment effects of the program (Buhl-Wiggers et al. 2020)

We use data from a five-year longitudinal RCT

- RCT was designed to study the NULP's impacts
  - Random sample of students tested using EGRA and followed across years
- We utilize two aspects of this study
  1. In 2013, 2016, and 2017 randomized students to teachers in schools with 2+ classrooms per grade (99% in 2013, 60% in 2016 & 2017).
  2. Schools randomized into control, full-cost program, and reduced-cost program.

We apply two restrictions to the data

1. **Two-teacher sample**: Only data from schools w/at least two teachers
   - To purge data of school effects, need to have at least two classrooms
   - We also require at least 5 students per teacher (& examine sensitivity to this cutoff)
2. **Longitudinal sample**: Teachers in the two-teacher sample observed in multiple years
   - To estimate teacher effects, need multiple observations per teacher

Also focus primarily on control-group teachers $\implies$ TVA under the *status quo*

Our sample includes over 1,300 teachers & almost 30,000 students

	NULP Evaluation Sample		Two-Teacher Sample		Longitudinal Sample	
	All Arms	Control Group	All Arms	Control Group	All Arms	Control Group
Schools	128	42	128	42	125	40
Teachers	1,382	470	1,096	365	475	146
with data on characteristics	878	281	871	282	435	132
Classrooms	2,200	728	1,763	568	1,138	347
Students sampled	27,943	8,948	27,608	8,820	24,217	7,468
Student-year obs	58,777	18,638	56,032	17,612	38,078	11,430

We have data on three test scores: Leblango (the local language), English, and Math. Show mainly Leblango, the focus of the NULP; some results for English too.

Estimating Classroom Effects

$$Y_{icgt} = \beta_0 + \beta_1 Y_{icg(t-1)} + \beta_2 X_{icgt} + \gamma_{cgt} + \zeta_g + \beta_3 Y_{icg(t-1)} * \zeta_g + u_{icgt}$$

Y_{icgt}: end-of-year test scores
$Y_{icg(t-1)}$: prior test scores for all three subjects
X_{icgt}: exogenous student controls
ζ_g: grade fixed effects
u_{icgt}: mean-zero error term

Estimating Classroom Effects

$$Y_{icgt} = \beta_0 + \beta_1 Y_{icg(t-1)} + \beta_2 X_{icgt} + \gamma_{cgt} + \zeta_g + \beta_3 Y_{icg(t-1)} * \zeta_g + u_{icgt}$$

Y_{icgt}: end-of-year test scores
$Y_{icg(t-1)}$: prior test scores for all three subjects
X_{icgt}: exogenous student controls
ζ_g: grade fixed effects
u_{icgt}: mean-zero error term

γ_{cgt}: classroom fixed effects (our coefficient of interest)

Estimating Classroom Effects

$$Y_{icgt} = \beta_0 + \beta_1 Y_{icg(t-1)} + \beta_2 X_{icgt} + \gamma_{cgt} + \zeta_g + \beta_3 Y_{icg(t-1)} * \zeta_g + u_{icgt}$$

Y_{icgt}: end-of-year test scores
$Y_{icg(t-1)}$: prior test scores for all three subjects
X_{icgt}: exogenous student controls
ζ_g: grade fixed effects
u_{icgt}: mean-zero error term

γ_{cgt}: classroom fixed effects (our coefficient of interest)
Focus on studying the SD of classroom effects, $\sqrt{Var(\gamma_{cgt})}$
Bootstrap SEs, clustered by school

Estimation Challenges

1. **Separating classroom effects from school effects.**

2. **Consistently estimating the variance of classroom effects.**

3. **Separating teacher effects from classroom effects.**

4. **Sorting of students into classrooms.**

Estimation Challenges

1. **Separating classroom effects from school effects.**
   - We re-scale classroom effects to be relative to the school mean (thus our estimates of the SD of classroom effects are lower bounds)

2. **Consistently estimating the variance of classroom effects.**

3. **Separating teacher effects from classroom effects.**

4. **Sorting of students into classrooms.**

Estimation Challenges

1. **Separating classroom effects from school effects.**
   - We re-scale classroom effects to be relative to the school mean (thus our estimates of the SD of classroom effects are lower bounds)

2. **Consistently estimating the variance of classroom effects.**
   - We analytically adjust the estimated variance for sampling error, following Araujo et al. (2016).

3. **Separating teacher effects from classroom effects.**

4. **Sorting of students into classrooms.**

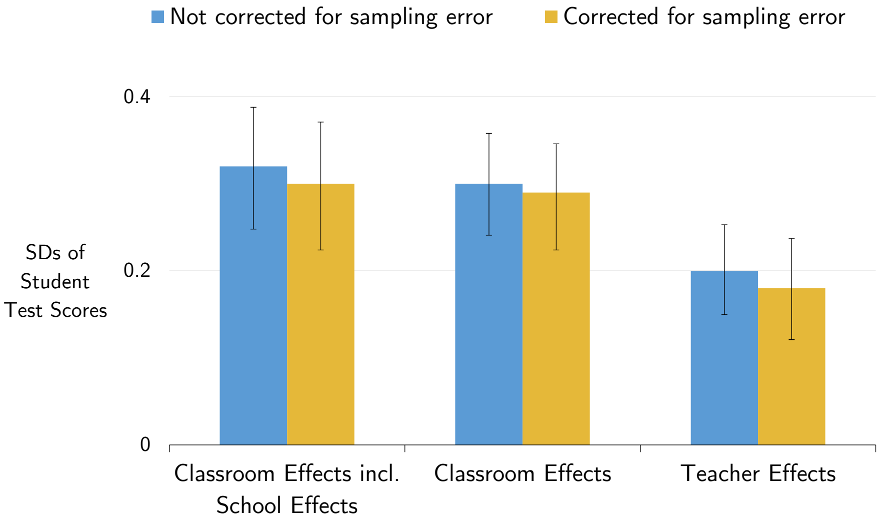
Estimation Challenges

1. **Separating classroom effects from school effects.**
   - We re-scale classroom effects to be relative to the school mean (thus our estimates of the SD of classroom effects are lower bounds)
2. **Consistently estimating the variance of classroom effects.**
   - We analytically adjust the estimated variance for sampling error, following Araujo et al. (2016).
3. **Separating teacher effects from classroom effects.**
   - We use the subset of teachers that appear in multiple years to estimate the stable component of teacher performance.
4. **Sorting of students into classrooms.**

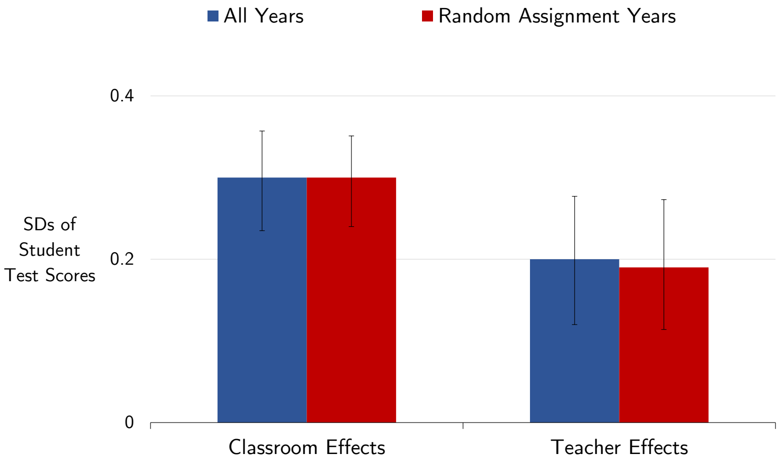
Estimation Challenges

1. **Separating classroom effects from school effects.**
   - We re-scale classroom effects to be relative to the school mean (thus our estimates of the SD of classroom effects are lower bounds)
2. **Consistently estimating the variance of classroom effects.**
   - We analytically adjust the estimated variance for sampling error, following Araujo et al. (2016).
3. **Separating teacher effects from classroom effects.**
   - We use the subset of teachers that appear in multiple years to estimate the stable component of teacher performance.
4. **Sorting of students into classrooms.**
   - We utilize the random assignment of children to teachers in 2013, 2016, and 2017 to assess the degree of bias present.

Addressing issues 1-3 shrinks the estimated SD of TVA

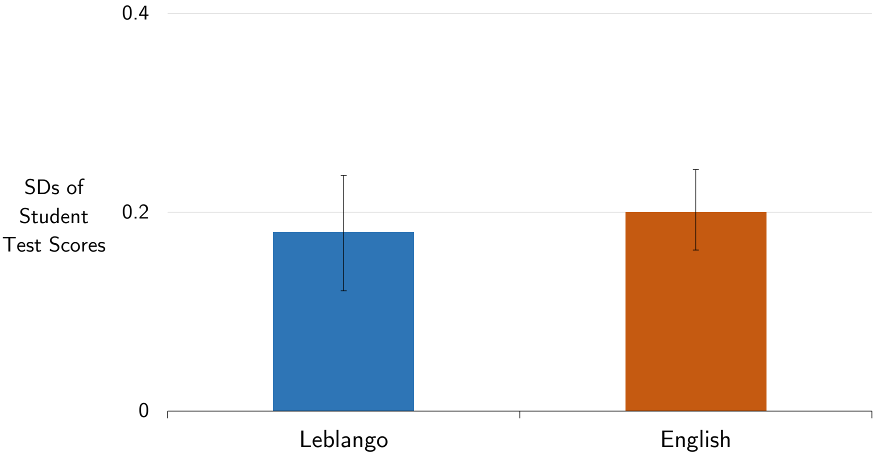


Random assignment $\Longrightarrow$ virtually unchanged TVA estimates



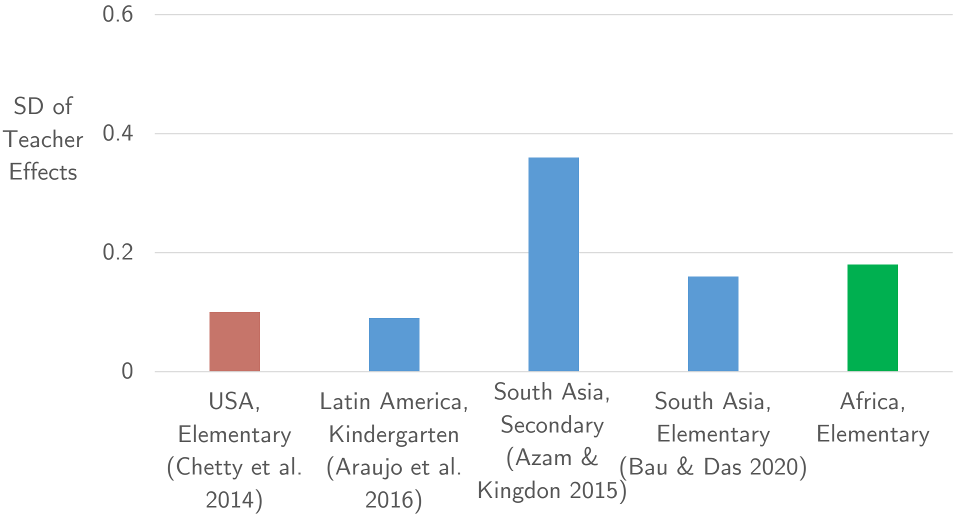
Other Robustness Checks

Spread of teacher effects is about the same for both languages

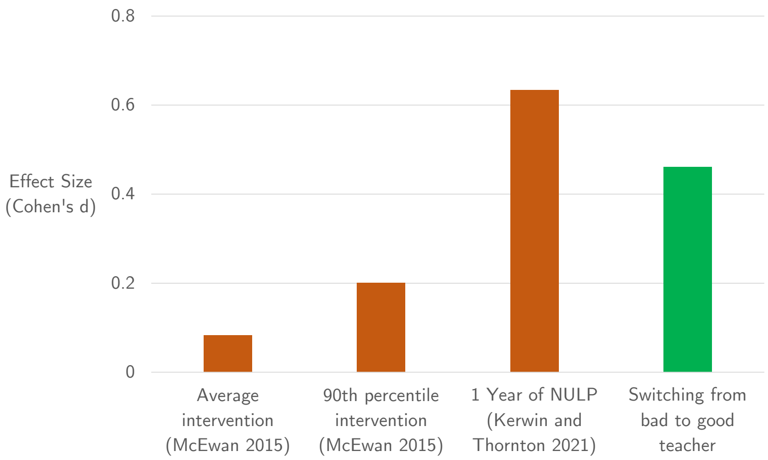


TVA is highly correlated across languages

Teaching quality matters even more in Africa than in the US



Changing from the 10th to the 90th percentile teacher helps as much as the very best education programs



Can we predict who is a good teacher?

Can we predict who is a good teacher?

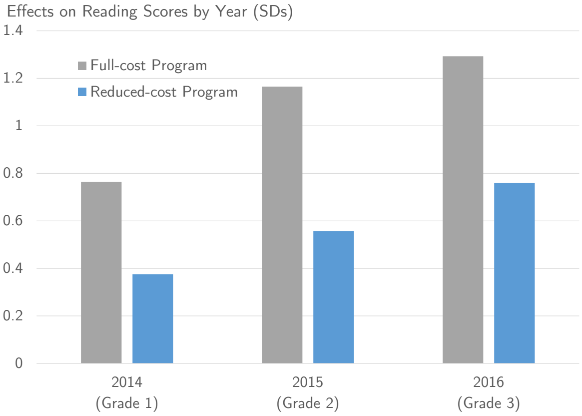
Not really.

Can we predict who is a good teacher?

Not really.

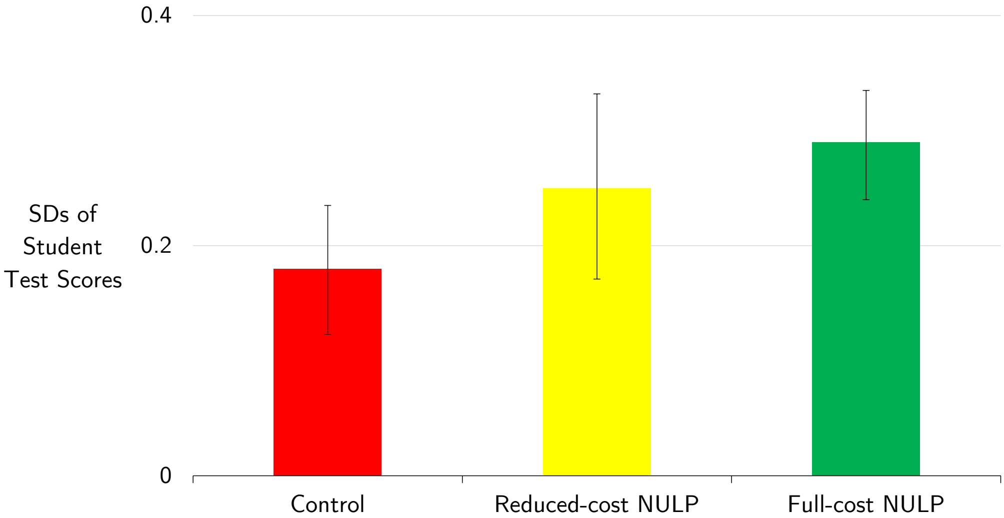
	Leblango TVA		English TVA	
	Classroom Effects	Teacher Effects	Classroom Effects	Teacher Effects
	(1)	(2)	(3)	(4)
$\geq$ Bachelor's Degree (1=Yes)	-0.075**	-0.051	-0.089*	-0.015
	(0.035)	(0.035)	(0.049)	(0.064)
Female (1=Yes)	-0.036	-0.004	-0.057	0.005
	(0.033)	(0.041)	(0.039)	(0.047)
$<$ 5 yrs of experience (1=Yes)	-0.006	0.064	0.073	0.250
	(0.140)	(0.235)	(0.136)	(0.272)
yrs of experience	-0.000	-0.001	0.006*	0.007
	(0.003)	(0.003)	(0.004)	(0.005)
$<$ 5 yrs of experience (1=Yes)$\times$	0.013	-0.013	-0.004	-0.032
yrs of experience	(0.045)	(0.073)	(0.045)	(0.092)
Observations	470	132	310	87
R-squared	0.017	0.014	0.034	0.048
Sample	Two-Teacher	Longitudinal	Two-Teacher	Longitudinal

The NULP program sharply improves test scores on average

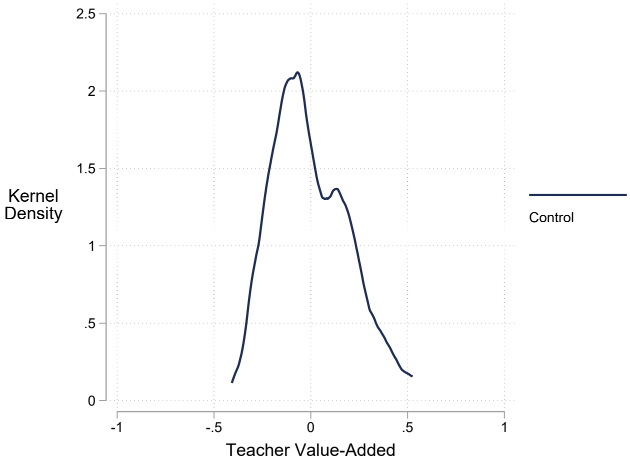


- P1 results for 2013 from Kerwin and Thornton (2021)
- P1-P3 results for 2014-2016 from Buhl-Wiggers et al. (2018, in progress)

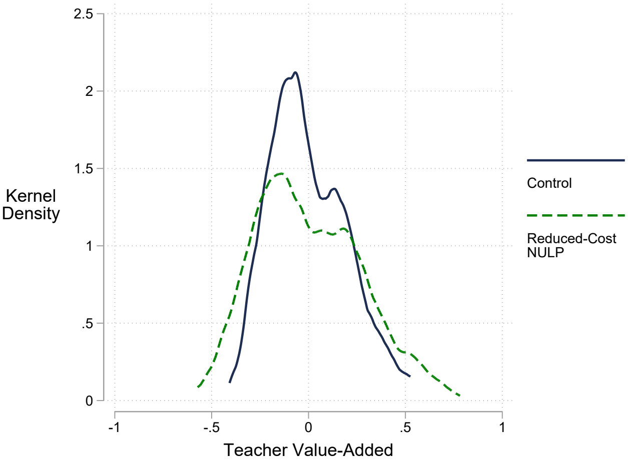
The NULP also increases the SD of teacher effects for Leblango



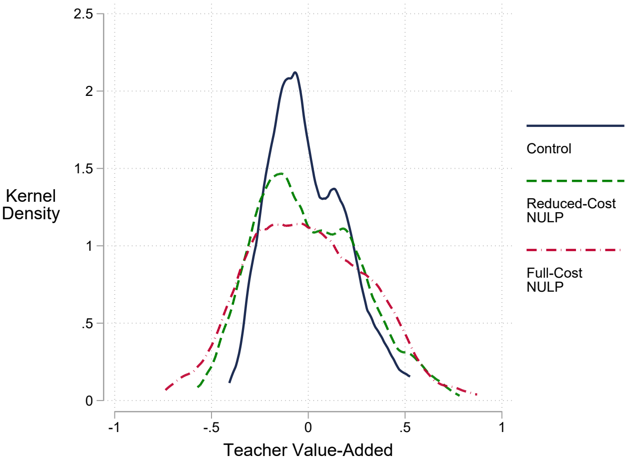
Spread of Leblango teacher effects is much wider in treatment arms



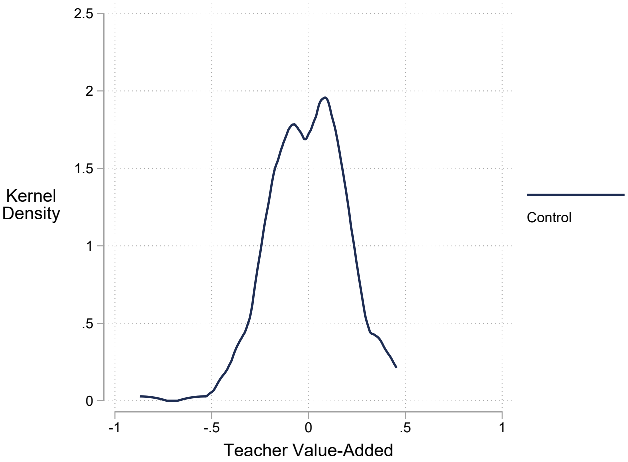
Spread of Leblango teacher effects is much wider in treatment arms



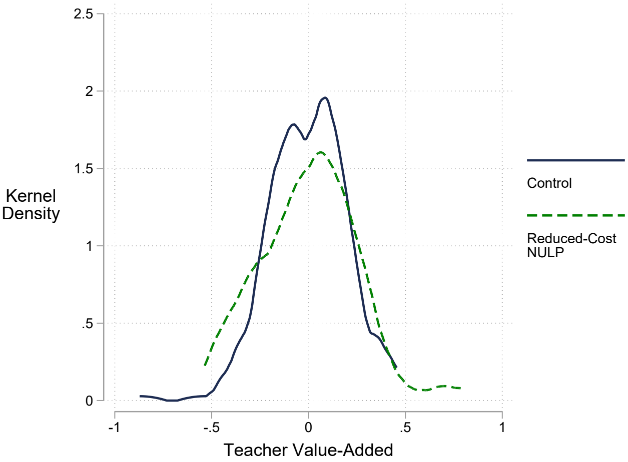
Spread of Leblango teacher effects is much wider in treatment arms



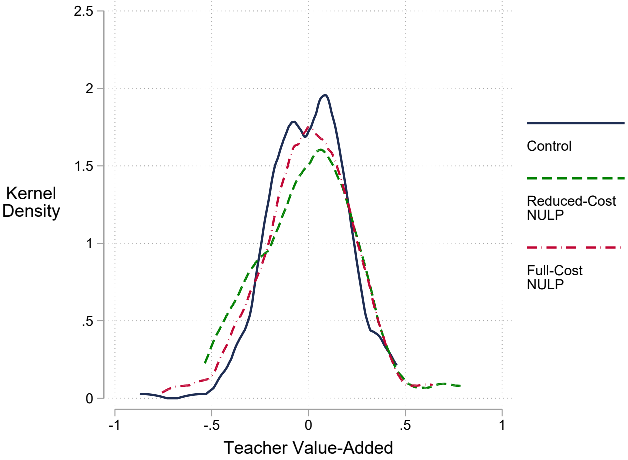
Smaller effects for English



Smaller effects for English



Smaller effects for English



The NULP intervention may have been rank-preserving

Partial test of rank preservation (Bitler et al. 2005): do fixed vars have same means in same quartile of TVA?

Not a high-powered test: Corr(TVA,Observables) is low

Across both subjects and both classroom + teacher effects (64 tests), reject null 7 times at 10% level—mostly for English teacher effects

	Difference between Full-Cost and Control in Teacher Characteristic:			
	Age	Gender	Experience	Schooling
	(1)	(2)	(3)	(4)
Leblango EGRA				
First quartile of TVA	-2.006	0.128	-1.558	0.025
	[-3.048,3.351]	[-0.204,0.202]	[-3.031,3.143]	[-0.123,0.122]
Second quartile of TVA	2.647	-0.008	-0.140	-0.082
	[-2.996,2.966]	[-0.216,0.205]	[-3.188,3.221]	[-0.201,0.174]
Third quartile of TVA	-0.753	-0.058	1.539	0.040
	[-3.759,4.043]	[-0.227,0.206]	[-3.424,3.392]	[-0.156,0.151]
Fourth quartile of TVA	-0.041	-0.087	-2.123	0.061
	[-3.068,3.058]	[-0.174,0.170]	[-3.184,3.192]	[-0.120,0.118]
Observations	284	291	281	291

How can we interpret these results?

- If NULP treatment were rank-preserving, then we could argue the gains are concentrated among strongest teachers
- If it is rank-inverting, conceivable that low-skill teachers gained a lot and high-skill teachers gained less
    - But this seems implausible
- Most likely: gains concentrated among best teachers, some amount of re-sorting due to treatment

Conclusion

- We present the first estimates of TVA for teachers in Africa

Conclusion

- We present the first estimates of TVA for teachers in Africa
  - Removing out school effects, estimation error, and bias due to sorting still imply that a 1 SD increase in teacher effectiveness increases student learning by about 0.2 SDs

Conclusion

- We present the first estimates of TVA for teachers in Africa
  - Removing out school effects, estimation error, and bias due to sorting still imply that a 1 SD increase in teacher effectiveness increases student learning by about 0.2 SDs
  - As in previous literature we find little correlation between TVA and teacher characteristics

Conclusion

- We present the first estimates of TVA for teachers in Africa
  - Removing out school effects, estimation error, and bias due to sorting still imply that a 1 SD increase in teacher effectiveness increases student learning by about 0.2 SDs
  - As in previous literature we find little correlation between TVA and teacher characteristics
  - Perhaps surprisingly, our conclusions are very similar to those for more resource-rich settings

Conclusion

- We present the first estimates of TVA for teachers in Africa
  - Removing out school effects, estimation error, and bias due to sorting still imply that a 1 SD increase in teacher effectiveness increases student learning by about 0.2 SDs
  - As in previous literature we find little correlation between TVA and teacher characteristics
  - Perhaps surprisingly, our conclusions are very similar to those for more resource-rich settings
- Taking the literature further, we examine what happens when we introduce a high-impact teacher training intervention

Conclusion

- We present the first estimates of TVA for teachers in Africa
  - Removing out school effects, estimation error, and bias due to sorting still imply that a 1 SD increase in teacher effectiveness increases student learning by about 0.2 SDs
  - As in previous literature we find little correlation between TVA and teacher characteristics
  - Perhaps surprisingly, our conclusions are very similar to those for more resource-rich settings
- Taking the literature further, we examine what happens when we introduce a high-impact teacher training intervention
  - Increases the spread of TVA in the local language
    - Probably by mainly helping good teachers and leaving bad ones behind
  - Has smaller effects on English, which was not targeted by the intervention
  - Somewhat discouraging implications for the hopes of teacher training programs

Thanks!

- If you have any other questions/comments/suggestions, please send them to me at jkerwin@umn.edu

Other Robustness Checks

- **Two different samples**: Classroom effects estimates are basically unchanged if estimated off the longitudinal sample instead of the entire two-teacher sample.

- **Small classes**: Main results set a minimum of 5 students per classroom. Very similar results for a 10-student minimum. 15-student minimum changes results, but only for local-language teacher effects.

- **Missing test scores**: Missing first-grade baseline scores for many students, and when present those scores are uninformative (mostly zero). Main estimates impute zero for all first-grade baseline scores. Dropping these obs instead barely affects results.

- **School-by-year effects**: Main analysis purges aggregate school effects; purging school-by-year effects instead makes little difference.

Back

TVA is highly correlated between English and Leblango

Correlations between teacher effect estimates across subjects, by study arm

Treatment doesn't change the correlations much

	Leblango TVA	English TVA
Control Group		
Leblango TVA	1	
English TVA	0.76	1
Reduced-cost NULP		
Leblango TVA	1	
English TVA	0.79	1
Full-cost NULP		
Leblango TVA	1	
English TVA	0.75	1

Back