Scaling up What Works:
Experimental Evidence on External Validity in Kenyan Education

Tessa Bold  Goethe University & IIES
Mwangi Kimenyi  Brookings Institution
Germano Mwabu  University of Nairobi
Alice Ng’ang’a  Strathmore University
Justin Sandefur  Center for Global Development

May 9, 2013
Contract teachers

- Muralidharan & Sundararaman (2008)
  Andhra Pradesh
  Contract teachers ⇒ +0.15 std. dev.

- Duflo, Dupas, & Kremer (2009)
  Western Kenya
  Contract teachers ⇒ +0.21 std. dev.
  Class size reduction ⇒ no effect on scores
Geography

Institutions

Millions of USD per annum spent on primary education

- MOE budget: $731
- Foreign aid to MOE: $16
- Foreign aid to NGOs: $3
Scale per se

Average TSC Salary Sh.19,400 ≈ $260 / month
Sh.10,000 ≈ $135 / month
Average PTA Salary Sh.4,200 ≈ $56 / month
Scale *per se*

Average TSC Salary  
Sh.19,400 $\approx$ $260 / month  
Sh.10,000 $\approx$ $135 / month

Average PTA Salary  
Sh.4,200 $\approx$ $56 / month

INCOME :- 10,000  
HOUSE HELP:- 7,000  
HOUSE RENT:- 7,000  
TRANSPORT :- 3,000  
FOOD 5,000  
SCHOOL FEES:- 10,000  
MISCELLANEOUS:- 5,000  
BALANCE:- NEGATIVE 25,000
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
Experimental Design

- Sampling
  - All 8 provinces, 14 (non-random) districts
  - Random sampling of schools w/ PTR > median

School-level randomization
- 192 schools
- 64 NGO, 64 Gov, 64 control

Intervention
- 1 add’l teacher per school
- Assigned to grade 2 or 3 in 2010
- 17 months exposure, immediate follow-up testing

Cross-cuts
- SMC training
- Central/local hiring
- High/low salary
Experimental Design

- **Sampling**
  - All 8 provinces, 14 (non-random) districts
  - Random sampling of schools w/ PTR > median

School-level randomization
- 192 schools
- 64 NGO, 64 Gov, 64 control

Intervention
- 1 add’l teacher per school
- Assigned to grade 2 or 3 in 2010
- 17 months exposure, immediate follow-up testing

Cross-cuts
- SMC training
- Central/local hiring
- High/low salary
Experimental Design

- **Sampling**
  - All 8 provinces, 14 (non-random) districts
  - Random sampling of schools with PTR > median

**School-level randomization**
- 192 schools
- 64 NGO, 64 Gov, 64 control

**Intervention**
- 1 add’l teacher per school
- Assigned to grade 2 or 3 in 2010
- 17 months exposure, immediate follow-up testing

**Cross-cuts**
- SMC training
- Central/local hiring
- High/low salary
Experimental Design

- **Sampling**
  - All 8 provinces, 14 (non-random) districts
  - Random sampling of schools w/ PTR > median

**School-level randomization**

- 192 schools
- 64 NGO, 64 Gov, 64 control

**Intervention**

- 1 add’l teacher per school
- Assigned to grade 2 or 3 in 2010
- 17 months exposure, immediate follow-up testing

**Cross-cuts**

- SMC training
- Central/local hiring
- High/low salary
Experimental Design

- **Sampling**
  - All 8 provinces, 14 (non-random) districts
  - Random sampling of schools with PTR $\geq$ median

School-level randomization

- 192 schools
- 64 NGO, 64 Gov, 64 control

**Intervention**

- 1 add’l teacher per school
- Assigned to grade 2 or 3 in 2010
- 17 months exposure, immediate follow-up testing

**Cross-cuts**

- SMC training
- Central/local hiring
- High/low salary
Project Timeline

- **Jul 2009**: Baseline evaluation for pilot
- **Aug 2009**: Union lawsuit
- **Jun 2010**: Pilot teachers placed in schools (NGO & Gov)
- **Oct 2010**: Gov hires 18,000 contract teachers
- **Sep 2011**: 18,000 made permanent
- **Oct 2011**: Final evaluation of pilot
Project Timeline

- **Jul 2009** Baseline evaluation for pilot
- **Aug 2009** Union lawsuit
- **Jun 2010** Pilot teachers placed in schools (NGO & Gov)
- **Oct 2010** Gov hires 18,000 contract teachers
- **Sep 2011** 18,000 made permanent
- **Oct 2011** Final evaluation of pilot
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
Treatment Effect of Contract Teachers on Test Scores
Experimental effects on teacher recruitment

Table: Labor supply of contract teachers

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>0.745</td>
<td>0.686</td>
<td>0.587</td>
</tr>
<tr>
<td></td>
<td>(0.034)***</td>
<td>(0.047)***</td>
<td>(0.064)***</td>
</tr>
<tr>
<td>NGO implementation</td>
<td>0.122</td>
<td>0.123</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.067)*</td>
<td>(0.065)*</td>
<td></td>
</tr>
<tr>
<td>High salary</td>
<td></td>
<td></td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.064)*</td>
</tr>
<tr>
<td>Local recruitment</td>
<td></td>
<td></td>
<td>0.143</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.065)**</td>
</tr>
<tr>
<td>Obs.</td>
<td>2,044</td>
<td>2,044</td>
<td>2,044</td>
</tr>
</tbody>
</table>
Treatment Effects

Table: \[ Y_{ijt} = \alpha_j + \beta Z_{jt} + \gamma (Z_{jt} \times Gov_{jt}) + \delta_t + \varepsilon_{ijt} \]

<table>
<thead>
<tr>
<th></th>
<th>ITT</th>
<th>LATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Z )</td>
<td>0.083</td>
<td>0.119</td>
</tr>
<tr>
<td>( T )</td>
<td>(0.076)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>NGO vs Gov:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Z )</td>
<td>0.180</td>
<td>0.245</td>
</tr>
<tr>
<td>( Z \times Gov )</td>
<td>(-0.197)</td>
<td>(-0.270)</td>
</tr>
<tr>
<td>( T )</td>
<td>(0.084)* *</td>
<td>(0.114)* *</td>
</tr>
<tr>
<td>( T \times Gov )</td>
<td>(-0.085)* *</td>
<td>(-0.122)* *</td>
</tr>
<tr>
<td>Obs.</td>
<td>14,975</td>
<td>14,975</td>
</tr>
</tbody>
</table>
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
### Mechanisms (1 of 2)

<table>
<thead>
<tr>
<th></th>
<th>Gov.</th>
<th>NGO</th>
<th>Difference</th>
<th>Corr. with value added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td><strong>Teacher characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>.379</td>
<td>.203</td>
<td>.177</td>
<td>-.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.075)**</td>
<td>(.092)</td>
</tr>
<tr>
<td>Post-secondary education</td>
<td>.138</td>
<td>.014</td>
<td>.124</td>
<td>-.131</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.045)**</td>
<td>(.149)</td>
</tr>
<tr>
<td>Advanced prof. qualification</td>
<td>.069</td>
<td>.095</td>
<td>-.026</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.043)</td>
<td>(.149)</td>
</tr>
<tr>
<td><strong>Local institutions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friend/relative of teacher</td>
<td>.667</td>
<td>.373</td>
<td>.294</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.100)**</td>
<td>(.100)</td>
</tr>
<tr>
<td>Presence</td>
<td>.628</td>
<td>.727</td>
<td>-.099</td>
<td>.101</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.110)</td>
<td>(.134)</td>
</tr>
<tr>
<td>Monitoring visit</td>
<td>.850</td>
<td>.961</td>
<td>-.111</td>
<td>.184</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.053)**</td>
<td>(.155)</td>
</tr>
<tr>
<td><strong>National politics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave. salary delay (months)</td>
<td>3.000</td>
<td>2.117</td>
<td>.883</td>
<td>-.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.291)**</td>
<td>(.034)*</td>
</tr>
<tr>
<td>Union represented me</td>
<td>.377</td>
<td>.149</td>
<td>.228</td>
<td>-.197</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.089)**</td>
<td>(.110)*</td>
</tr>
<tr>
<td>Took union action</td>
<td>.533</td>
<td>.471</td>
<td>.063</td>
<td>-.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.096)</td>
<td>(.097)</td>
</tr>
</tbody>
</table>
## Mechanisms (2 of 2)

<table>
<thead>
<tr>
<th></th>
<th>Union identification</th>
<th>Test-score gains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$Z \times \text{Gov}$</td>
<td>0.084</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>$Z \times \text{NGO} \times \text{Union exposure}$</td>
<td>0.083</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.183)</td>
</tr>
<tr>
<td>$Z \times \text{Gov} \times \text{Union exposure}$</td>
<td>0.548***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.168)</td>
<td></td>
</tr>
<tr>
<td>$Z \times \text{NGO} \times \text{Exposure to gov't scale-up}$</td>
<td></td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.115)</td>
</tr>
<tr>
<td>$Z \times \text{Gov} \times \text{Exposure to gov't scale-up}$</td>
<td></td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.154)</td>
</tr>
<tr>
<td>Observations</td>
<td>100</td>
<td>95</td>
</tr>
</tbody>
</table>
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
Heterogeneity

PTR

Geographic density

Baseline Test Scores
Heterogeneity

PTR

Geographic density

Baseline Test Scores
Heterogeneous treatment effects

Does impact vary across following dimensions? (overall, and for Gov and NGO individually)

- Geographic remoteness
- Initial pupil-teacher ratio
- Initial test scores

Western baseline scores 1/2 S.D. below mean ⇒ Gov-NGO gap 0.05 S.D. narrower in Western
Heterogeneous treatment effects

Does impact vary across following dimensions? (overall, and for Gov and NGO individually)

- Geographic remoteness
- Initial pupil-teacher ratio
- Initial test scores (—) only in Gov sample

Western baseline scores 1/2 S.D. below mean ⇒ Gov-NGO gap 0.05 S.D. narrower in Western
Heterogeneous treatment effects

Does impact vary across following dimensions? (overall, and for Gov and NGO individually)

- Geographic remoteness X
- Initial pupil-teacher ratio X
- Initial test scores (−) only in Gov sample

Western baseline scores 1/2 S.D. below mean
⇒ Gov-NGO gap 0.05 S.D. narrower in Western
Outline

Experimental design & context

Institutions
  Horse race
  Mechanisms

Geography

Conclusion
Conclusions (1 of 2)

- Geography & heterogeneous response
  - Intervention is progressive
  - But little reason question external validity from Western Kenya

- Institutions & partner selection bias
  - Horse race results: Institutions matter
    - e.g., local nepotism in gov’t sector ↓ of scale

- Scale & see-saw effects
  - Hint that gov’t failure was a function of scale
    - e.g., union affiliation, salary delays
Conclusions (2 of 2)

- Lessons for impact evaluation
  - Is critique of external validity externally valid?
  - External validity vs. construct validity
  - Problem of IE not RCTs
  - NGOs as a laboratory vs. an accountability system
## Compliance & Contamination

### Table:

<table>
<thead>
<tr>
<th>Compliance</th>
<th>All Schools</th>
<th>Treated</th>
<th>Control</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class size</td>
<td></td>
<td>60.229</td>
<td>69.047</td>
<td>-8.818</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.179)**</td>
<td>(5.919)**</td>
<td>(6.131)</td>
</tr>
<tr>
<td>Teacher ever in correct class</td>
<td></td>
<td>.953</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.020)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher always in correct class</td>
<td></td>
<td>.729</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.043)**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Contamination

<table>
<thead>
<tr>
<th>Contamination</th>
<th>All Schools</th>
<th>Treated</th>
<th>Control</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log enrollment in treatment cohort</td>
<td></td>
<td>4.954</td>
<td>5.036</td>
<td>-.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.064)**</td>
<td>(.074)**</td>
<td>(.103)</td>
</tr>
<tr>
<td>Change in log cohort enrollment</td>
<td></td>
<td>-.109</td>
<td>-.093</td>
<td>-.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.023)**</td>
<td>(.035)**</td>
<td>(.040)</td>
</tr>
<tr>
<td>No. of teachers from 18,000 program</td>
<td></td>
<td>.667</td>
<td>.500</td>
<td>.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.107)**</td>
<td>(.135)**</td>
<td>(.177)</td>
</tr>
</tbody>
</table>
## Compliance & Contamination

### Table:

<table>
<thead>
<tr>
<th></th>
<th>MOE</th>
<th>NGO</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compliance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class size</td>
<td>60.470</td>
<td>59.980</td>
<td>.490</td>
</tr>
<tr>
<td></td>
<td>(5.001)**</td>
<td>(3.687)**</td>
<td>(6.131)</td>
</tr>
<tr>
<td>Teacher ever in correct class</td>
<td>.966 (.024)**</td>
<td>.938 (.035)**</td>
<td>.029 (.042)**</td>
</tr>
<tr>
<td>Teacher always in correct class</td>
<td>.763 (.058)**</td>
<td>.688 (.072)**</td>
<td>.075 (.092)**</td>
</tr>
<tr>
<td><strong>Contamination</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log enrollment in treatment cohort</td>
<td>4.951 (.070)**</td>
<td>4.957 (.105)**</td>
<td>-.007 (.094)**</td>
</tr>
<tr>
<td>Change in log cohort enrollment</td>
<td>-.137 (.028)**</td>
<td>-.079 (.035)**</td>
<td>-.059 (.037)**</td>
</tr>
<tr>
<td>No. of teachers from 18,000 program</td>
<td>.727 (.163)**</td>
<td>.607 (.140)**</td>
<td>.175 (.189)**</td>
</tr>
</tbody>
</table>