Three levels of analysis from large-scale research on interventions to improve literacy in South Africa

Quantitative RCT findings: UKFIET, September 2017
Cas Prinsloo and Nompumelelo Mohohlwane
PRESENTATION OUTLINE

• Locating impact evaluation in SA education research
• The Randomised Control Trial (RCT) design
• Year 2 impact evaluation results
  – Overall impacts on reading outcomes
  – For whom and where did it work?
  – Cost-effectiveness analysis
Locating impact evaluation in SA education research

• Qualitative work

• Systemic analysis with mixed methods
  – Taylor, Vinjevold, Muller (2003); Fleisch (2008)

• Descriptive quantitative work
  – Reddy (2006); Taylor & Yu (2008); Spaull (2011)

• Correlational analysis
  – Crouch & Mabogoane (1998); Van der Berg (2008); Gustafsson (2007); Spaull (2012); Shepherd (2011)

• Moving toward causal quantitative analysis
Why evaluate using an RCT with mixed methods?

• The evaluation problem:
  – We cannot observe the counterfactual:
  – 2 alternative scenarios for the same person or group.

• So we have to identify or construct comparison groups to estimate the counterfactual.

• The big question is: when is a comparison group a valid estimate of the counterfactual?
  – Internal validity

• Selection bias
  – Libraries and learning outcomes
Why an RCT?
Why an RCT?
230 schools in North West

Random assignment creates a plausible estimate of the counterfactual

<table>
<thead>
<tr>
<th>Control group</th>
<th>“Training”</th>
<th>“Coaching”</th>
<th>“Parent Involvement”</th>
</tr>
</thead>
<tbody>
<tr>
<td>(80 schools)</td>
<td>Lesson plans, reading materials + central teacher training (50 schools)</td>
<td>Lesson plans, reading materials + on-site coaching (50 schools)</td>
<td>(50 schools)</td>
</tr>
</tbody>
</table>
Critiques of RCTs

• Necessary and sufficient conditions for impact evaluations (internal and external validity)
• Internal validity = causal inference
• External validity = transferability to population
  – Context: geography, time, etc...?
    • E.g. Private schools, class size
  – Special experimental conditions
    • Hawthorne effects
    • Implementation agent
    • System support
• Mechanisms of change: How? Why?
Critiques of RCTs: What to do?

• Choose a representative & relevant study population
• Investigate sub-group effects
• Investigate intermediate outcomes
• Incorporate mixed methods research
EGRS design
458 registered primary schools with enrolments in grades 1-4

Apply a series of exclusions

- Exclude schools not using Setswana as language of instruction
- Exclude small schools and large schools
- Exclude schools with missing ANA data
- Exclude schools with missing ANA data
- Exclude affluent schools (quintiles 4 and 5)
- Exclude 8 pilot schools
- Exclude replacement schools
- Exclude problem schools identified by PED

Sampling Frame of 230 schools

Create 10 strata by school size, school socio-economic status and ANA performance

Randomly assign schools within each stratum to T1, T2, T3 and Control

This yields 4 treatment groups

| T1: Teacher training (50 schools) | T2: Coaching (50 schools) | T3: Parent involvement (50 schools) | Control group (80 schools) |
3 waves of data collection

• Wave 1: “Baseline”
  – Start of Grade 1, Feb 2015
  – Randomly sampled 20 learners per school

• Wave 2: “Midline”
  – End of Grade 1, Oct/Nov 2015

• Wave 3: “Endline”
  – End of Grade 2, Oct/Nov 2016
  – Included those repeating grade 1
**Attrition & repetition**

### Attrition
- Due to leaving school or to absenteeism
- Related to gender, poverty and being in a specific district, but not to learner performance

### Repeating Grade 1
- Strongly related to grade 1 reading achievement, to being in a specific district and to gender.

### In Grade 2
- 3726 pupils remain in the sample
Secondary benefits of longitudinal data

Proportion repeating grade 1 in 2016

Decile of achievement at end of grade 1 in 2015

Boys
Girls
Main results: Multi-variable regression model

- Reading = function of:
  - Baseline achievement
  - Gender
  - Age
  - District
  - Community SES
  - Intervention group
Main results: Impact including repeaters

- **Training**: Estimated effect size (SD) = 0.11
- **Coaching**: Estimated effect size (SD) = 0.25
- **Parents**: Estimated effect size (SD) = 0.1

About 30% of a year of learning
Main results: Impact for those with 2 years of interventions

About 40% of a year of learning
• How large are these impacts?
  —Relative to a year of learning?
29% (including repeaters)
39% (excluding repeaters)
Who benefits most from the interventions?

- Boys catch up to some extent
- Large-classes benefited most
- Middle-to-top performing learners benefited most
- Impact concentrated in urban schools
Boys catch up to some extent

[Graph showing percentage achievement at different levels of paragraph reading words per minute for Control Males, Coaching Males, Control Females, and Coaching Females.]
Boys catch up to some extent

![Graph showing percentage of students achieving at least this level in paragraph reading words per minute, with lines for Coaching Females and Coaching Males.]
Boys catch up to some extent
Large(ish)-classes benefited most

**Training**
- Estimated effect size (SD): -0.079

**Coaching**
- Estimated effect size (SD): 1

- Tercile 1
- Tercile 2
- Tercile 3
Middle-to-top performing learners benefited most
Cost-effectiveness analysis

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Coaching</th>
<th>Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total annual costs per 50 schools</td>
<td>R2.34m</td>
<td>R3m</td>
<td>R0.96m</td>
</tr>
<tr>
<td>Per learner annual cost</td>
<td>R 626</td>
<td>R 804</td>
<td>R 256</td>
</tr>
<tr>
<td>Estimated impact</td>
<td>0.11 SD</td>
<td>0.25 SD</td>
<td>0.1 SD</td>
</tr>
<tr>
<td>Test score gains per R1000 spent</td>
<td>0.18 SD</td>
<td>0.31 SD</td>
<td>0.38 SD</td>
</tr>
</tbody>
</table>
Cost-effectiveness analysis

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Coaching</th>
<th>Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total annual costs per 50 schools</td>
<td>R2.34m</td>
<td>R3m</td>
<td>R0.96m</td>
</tr>
<tr>
<td>Per learner annual cost</td>
<td>R 626</td>
<td>R 804</td>
<td>R 256</td>
</tr>
<tr>
<td>Estimated impact</td>
<td>0.11 SD</td>
<td>0.25 SD</td>
<td>0.1 SD</td>
</tr>
<tr>
<td>Test score gains per R1000 spent</td>
<td>0.18 SD</td>
<td>0.31 SD</td>
<td>0.38 SD</td>
</tr>
</tbody>
</table>
Cost-effectiveness analysis

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Coaching</th>
<th>Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total annual costs per 50 schools</td>
<td>R2.34m</td>
<td>R3m</td>
<td>R0.96m</td>
</tr>
<tr>
<td>Per learner annual cost</td>
<td>R 626</td>
<td>R 804</td>
<td>R 256</td>
</tr>
<tr>
<td>Estimated impact</td>
<td>0.11 SD</td>
<td>0.25 SD</td>
<td>0.1 SD</td>
</tr>
<tr>
<td>Test score gains per R1000 spent</td>
<td>0.18 SD</td>
<td>0.31 SD</td>
<td>0.38 SD</td>
</tr>
</tbody>
</table>
Cost-effectiveness analysis

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Coaching</th>
<th>Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total annual costs per 50 schools</td>
<td>R2.34m</td>
<td>R3m</td>
<td>R0.96m</td>
</tr>
<tr>
<td>Per learner annual cost</td>
<td>R 626</td>
<td>R 804</td>
<td>R 256</td>
</tr>
<tr>
<td>Estimated impact</td>
<td>0.11 SD</td>
<td>0.25 SD</td>
<td>0.1 SD</td>
</tr>
<tr>
<td>Test score gains per R1000 spent</td>
<td>0.18 SD</td>
<td>0.31 SD</td>
<td>0.38 SD</td>
</tr>
</tbody>
</table>
Conclusions

• 230 schools data:
  – Shows the magnitudes of impact
  – Suggests where programmes worked best or not at all
  – Provides an indication of intermediate outcomes (why, how)

• 60-school lesson observation
  – More conclusive evidence on classroom practice

• Case studies
  – Quality of practice
  – Indication of what prevented impact in rural areas
Thank you!

www.education.gov.za
facebook: DBE SA
twitter: @DBE_SA
callcentre@dbe.gov.za
callcentre: 0800 202 933
White’s (2009) six principles of Theory-Based Evaluation

1. Map out the causal chain (programme theory)
2. Understand context
3. Anticipate sub-group effects and have a large enough sample to disaggregate on sub-groups
4. Rigorous evaluation of impact using a credible counterfactual
5. Rigorous factual analysis of links in the causal chain
6. Combine quantitative & qualitative methods