Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments

Patrick McEwan
Wellesley College

pmcewan@wellesley.edu
Paper available at www.patrickmcewan.net
Why a meta-analysis of RCTs?

• Meta-analysis summarizes evidence from multiple impact evaluations, for diverse treatments and settings
  – Average effect sizes and correlates of effect sizes
• Randomized, controlled trials
  – Gold standard for answering causal questions
  – Growing number applied in poorer countries
• What patterns do RCTs reveal? What “works” for improving learning?
External validity

• Do causal effects “[hold] over variations in persons, settings, treatment variables, and measurement variables”? (Shadish, Cook, & Campbell, 2002, p. 507).

• Some approaches
  1. Formal sampling of persons, settings, etc.
  2. Causal explanation, a.k.a. “inside the black box”
  3. Ruling out irrelevancies, making discriminations
Varities of treatments

**Instructional**
- Teacher training in pedagogy and content; mentoring/coaching
- Textbooks and other materials
- Computers, software, and technology
- Size of learning groups (class size reduction, small-groups within classes)
- Composition of learning groups

**Health and nutrition**
- Micronutrients
- School feeding programs
- Treatment or prevention of
  - Intestinal helminths
  - Malaria
- Vision screening and other health services

**Incentives**
- Information for students, parents or schools ("report cards")
- Performance incentives for students or teachers
- Teachers with flexible contracts
- School-based management
RCTs yield net policy effects
Literature search

• Conducted August 2012 to February 2013
• Sources of RCTs
  – Reviews by economists (e.g., Glewe et al. 2011; Petrosino et al. 2012, Kremer & Holla, 2009; Kremer et al., 2013)
  – Reviews in health (meta-analysis of de-worming, iron/micronutrient supplementation, school feeding, malaria)
  – Keyword search on websites: J-PAL, CEGA, IPA, IADB, World Bank, RTI, Stanford REAP, NBER working papers
  – Hand-search of selected journals
Criteria for study inclusion

• 76 RCTs evaluating 110 treatments
  1. Lower to upper-middle income country
  2. “Primary” (grades 1-8), ages 6-14 if grades not reported
  3. Random assignment of children (or clusters of children) to education/health intervention in school setting (vs. “business-as-usual”)
  4. ≥1 continuously-measured learning outcome in language/reading, math, or composite
  5. Sufficient data to calculate effect sizes and standard error in full sample
Coding of study attributes

• Filemaker database
• Six tables: experiments, papers, treatment arms, follow-ups, outcomes, effects
• Experiment defined as ≥1 treatment arm(s) and single control group
  – Modal experiment is 1 treatment arm and control, reported in single paper
  – Some experiments reported in multiple papers
    • Muralidharan & Sundararaman (2010a, 2011); Das et al. (2011)
  – Some papers report multiple experiments
    • Banerjee et al. (2007), Pradhan et al. (2011)
• Recent growth in RCTs
• Mostly 1-year experiments
• Only a few with longer-run follow-ups
• Kenya and India are popular research sites
• China increasingly so (Stanford’s REAP)
• Latin America under-represented

Note: Numerical codes identify experiments in McEwan (2013). Solid circles indicate the date of baseline data collection in each experiment (defined as a control group and one or more treatment arms). Hollow circles indicate follow-up data collection(s), and lines indicate the duration of the treatment(s), in months.
Analysis of effect sizes

- Random effects meta-analysis

\[ \hat{\theta}_{ijk} = \theta + e_{ijk} + u_{ijk}, e_{ijk} + u_{ijk} \sim N(0, \nu_{ijk} + \sigma^2_\theta). \]

- Theta-hat is \( i \)th ES estimate in experiment \( j \), clustered in study \( k \) (same “study” if overlapping samples or identical treatments).

- \( e \) is ES estimation error, \( u \) is error due to unobserved features of treatment, sample, study quality, etc.

- Estimate via weighted least squares (Ringquist 2013)
  - Inverse variance weights
  - Further weight by \( 1/n \) where \( n \) is # of effect sizes in experiment-by-treatment arm cell

- Cluster standard errors by study
## Summary of averages within category

<table>
<thead>
<tr>
<th></th>
<th>Mean effect size</th>
<th>p-value</th>
<th>Number of:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Studies</td>
<td>Experiments</td>
</tr>
<tr>
<td><strong>Instructional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computers or technology</td>
<td>0.150</td>
<td>&lt;0.001</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Teacher training</td>
<td>0.123</td>
<td>&lt;0.001</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>Class size or composition</td>
<td>0.117</td>
<td>0.018</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Instructional materials</td>
<td>0.078</td>
<td>&lt;0.001</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Monetary grants</td>
<td>-0.011</td>
<td>0.723</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>Health or nutrition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food, beverages, micronutrients</td>
<td>0.035</td>
<td>0.054</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>De-worming drugs</td>
<td>0.013</td>
<td>0.388</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td><strong>Incentives</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contract/volunteer teachers</td>
<td>0.101</td>
<td>&lt;0.001</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Student/teacher performance incentives</td>
<td>0.095</td>
<td>0.066</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>School management/supervision</td>
<td>0.055</td>
<td>0.168</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Informational treatments</td>
<td>0.049</td>
<td>0.240</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>
Example: computers and technology

Notes: Numerical codes identify experiments in McEwan (2013). Diamonds and brackets indicate effect sizes and 95% confidence intervals. P-value of mean effect size is calculated using wild cluster bootstrap-t.
Example: food, micronutrients

Notes: Numerical codes identify experiments in McEwan (2013). Diamonds and brackets indicate effect sizes and 95% confidence intervals. P-value of mean effect size is calculated using wild cluster bootstrap-t.
Caveats of descriptive analysis

• Categories *not* mutually exclusive
  – Instructional materials alone seem to have few effects
  – Contract teacher treatments conflated with class size reduction

• Meta-regressions
  – Control for moderators: implementer, exposure to treatment, context, outcome measures, study quality
  – Pattern of results does not change much if we control for these variables in meta-regression
Issue 1: non-learning outcomes

• School enrollment, attendance, and attainment
  – De-worming and school feeding programs have stronger attainment effects, but weak learning effects
  – CCTs have strong attainment effects, and mixed learning effects (e.g., Mexico vs. Nicaragua)

• Suggestive evidence that additional time in school is not always used productively

• Few factorial designs to jointly evaluate “supply-side” instructional interventions alongside CCTs or health (PRAF tried in 2000!)
Issue 2: Cost-effectiveness analysis

• Cost-benefit analysis would be desirable; not likely in near term

• CEA instead? Issues
  – Mixed cost data (56% of treatment reported no costs; others reported some). Uncertain methods and data.
  – Multiple learning outcomes

• Exercise: use consistent J-PAL cost estimates for subset of meta-analytic sample
  – Some modifications: 20% deadweight loss, no transfers, modified to match T-C contrast in sample
Social cost per 0.2 s.d.

Note: The x-axis indicates the cost per 0.2 units of the effect size on a log scale. Hollow diamonds indicate that the underlying effect size is not statistically significant at 10%. Costs per students were calculated with J-PAL data. See McEwan (2013) for details.
Issues 3: power and sample size

• Minimum detectable effect size (MDES) of cluster-randomized trial
  – 100 schools split across T/C, 50 students/school, ICC=0.2
  – MDES=0.26 (mainly affected by # of clusters and ICC)

• Issues
  – 0.26 is large relative to “typical” effect sizes in many categories
  – ICC may be under-stated (but little evidence)
    • PIRLS 4th grade reading: Colombia=0.44, Honduras=0.35...
  – Vital role of sensible stratification, good baseline to reduce MDES
Issue 4: external validity

• Meta-regression is one approach to assessing external validity

• Experiments could build in others
  – Random rather than convenience sample
  – Rigorous subgroup analysis (no “fishing” expeditions)
  – Factorial designs
  – Implementation and process-related data
  – Good outcome measures

• And thoughtfully-conceived quasi-experiments
Issue 5: reporting of experimental results

• Common missing information in reports
  – Sample at baseline and follow-up; attrition
  – Complete set of statistical models
    • No controls for strata, or only “kitchen-sink” model reported
  – Complete description of treatment and outcome measures
  – Costs

• Guidelines needed? (e.g., CONSORT)
Take-aways

• De-worming, meals, grants are not silver bullet for learning
• Smaller mean effects for diffuse categories of management/supervision and information
• Larger effects for computers, training, materials, class size/composition, teacher/student incentives, contract teachers
  – But materials alone do not seem to increase learning
  – Contract teacher interventions conflated with class size, not robust in meta-regression, but data are still thin
• Cost caveat
  – Low costs could make some smaller-ES interventions much more attractive (information?)
  – High costs could make some larger-ES interventions less so (CAI?)
Wish list

- Better data and reporting of data on costs, implementation, and causal mechanisms
- More partnerships with governments
- Longer-run follow-ups
- Focus on multiple, valid outcomes
  - education, health, labor market
- Factorial experiments, with many-layered treatments
  - Health and instructional inventions
  - Incentive-based interventions and instructional interventions
Thank you!

- Please visit [www.patrickmcewan.net](http://www.patrickmcewan.net) for paper
Monetary grants

Mean = -.011
p-value = .723
De-worming drugs

Figure 10: Treatments with de-worming drugs

Notes: Numerical codes identify experiments (see the Appendix). Diamonds and brackets indicate effect sizes and 95% confidence intervals. See text for details.

Mean=.013
p-value=.388
Information (report cards, etc.)

Mean = 0.049
p-value = 0.24
School management/supervision

Notes: Numerical codes identify experiments (see the Appendix).

Diamonds and brackets indicate effect sizes and 95% confidence intervals. See text for details.

- 69, Philippines, L, 1, "Multi-level materials", parent-teacher partnerships
- 69, Philippines, L, 1, School feeding, parent-teacher partnerships
- 69, Philippines, M, 1, School feeding, parent-teacher partnerships
- 52, Kenya, L, 1, Class size reduction, contract teacher, school-based management
- 52, Kenya, M, 1, Class size reduction, contract teacher, school-based management
- 69, Philippines, M, 1, "Multi-level materials", parent-teacher partnerships
- 39, Indonesia, L, 1, Linkage between school committee and village, election of school committee
- 52, Kenya, M, 1, Class size reduction, school-based management
- 52, Kenya, L, 1, Class size reduction, school-based management
- 37, Indonesia, L, 1, Linkage between school committee and village
- 39, Indonesia, M, 1, Linkage between school committee and village, election of school committee
- 36, Indonesia, L, 1, Election of school committee
- 36, Indonesia, L, 2, Class size reduction, school-based management
- 38, Indonesia, L, 1, Training of school committee
- 40, Indonesia, L, 1, Linkage between school committee and village, training of school committee
- 59, Madagascar, C, 1, School report cards, district, sub-district, and school management training
- 52, Kenya, M, 2, Class size reduction, contract teacher, school-based management
- 59, Madagascar, C, 2, School report cards, district, sub-district, and school management training
- 37, Indonesia, M, 1, Linkage between school committee and village
- 39, Indonesia, M, 1, Linkage between school committee and village, election of school committee
- 36, Indonesia, L, 1, Election of school committee
- 40, Indonesia, M, 1, Linkage between school committee and village, training of school committee
- 59, Madagascar, C, 2, School report cards, district and sub-district management training
- 41, Indonesia, L, 1, Training of school committee, election of school committee
- 36, Indonesia, M, 1, Election of school committee
- 59, Madagascar, C, 2, School report cards, district management training
- 38, Indonesia, M, 1, Training of school committee
- 41, Indonesia, M, 1, Training of school committee, election of school committee
- 18, Gambia, L, 3, School management training, grants
- 59, Madagascar, C, 1, School report cards, district and sub-district management training
- 38, Indonesia, L, 1, Training of school committee
- 59, Madagascar, C, 1, School report cards, district management training
- 18, Gambia, M, 3, School management training, grants

Mean = 0.055
p-value = 0.168
Teacher or student incentives

Figure 12: Treatments with performance incentives

Notes: Numerical codes identify experiments (see the Appendix). Diamonds and brackets indicate effect sizes and 95% confidence intervals. See text for details.
Contract/volunteer teachers

Figure 13: Treatments with contract or volunteer teachers

Notes: Numerical codes identify experiments (see the Appendix). Diamonds and brackets indicate effect sizes and 95% confidence intervals. See text for details.

Mean = 0.101
p-value < 0.001
Class size/composition

Mean = .117

p-value = .018