The Contribution of Systematic Reviews to Understanding School Effectiveness.

Philip Davies
International Initiative for Impact Evaluation [3ie]
Why Do We Need Systematic Reviews of Evidence

• Sheer amount and flow of information/research
• Variable quality of research outputs
• Need to separate the wheat from the chaff
• Problems of publication bias
• Need for the balance of evidence
• Limitations of single studies
Limitations of Single Studies

- Single studies can misrepresent the balance of research evidence
- Illuminate only one part of a policy issue
- Sample-specific
- Time-specific
- Context-specific
- Often of poor methodological quality
- Consequently, biased
Systematic Reviews

• “Attempt to discover the consistencies and account for the variability in similar-appearing studies”

• “Seeking generalisations also involves seeking the limits and modifiers of generalisations”

• Identify the contextual-specificity of available research and evidence”

(Cooper and Hedges, 1994:4).
Types of Research Synthesis

- Statistical Meta-Analyses (6-18 months)
- Narrative Systematic Reviews (6-12 Months)
- Rapid Evidence Assessments (1-3 Months)
- Evidence Maps and Gap Maps (1 Month)
- Meta-Ethnography/Qualitative Synthesis (6-12 Months)
What works in developing nations to get children into school and keep them there: A systematic review of experimental and quasi-experimental evaluations

A REPORT FUNDED BY THE INTERNATIONAL INITIATIVE FOR IMPACT EVALUATION (3IE)

DECEMBER 2011
Background

• Education is critical to economic development and social welfare particularly in economically developing countries

• Many interventions to increase school attendance and to improve quality of education in developing countries.

But:

• *No systematic review of the evidence*
Meta-Analytical Reviews

• Involves data-pooling and statistical synthesis of independent studies
• And aggregating/cumulating samples and findings
• Seeks to measure and control bias
Objectives

• To determine the *effects of interventions* implemented in developing countries as measured by *students’ enrollment, attendance, graduation, and progression*.

• To determine the *effects of interventions* on *learning outcomes* as measured by students’ test scores, grades, and other achievement measures.
Methods – Inclusion Criteria

Studies that:

1. Assess the impact of an intervention that included *primary or secondary school outcomes* (Kindergarten-12th grade in the U.S. context) relevant to the primary research question;

2. Use a *randomized controlled trial*, or a *quasi-experimental* approach in baseline control on primary outcome was included;

3. Be conducted in a *country classified* as a “low or middle income nation” by the World Bank at the time the intervention being studied was implemented;

4. Include at least *one quantifiable primary outcome measure* (enrollment, attendance, dropout, or progression);

5. Be published or made available before *December 2009*, without regard to language or publication type; and

6. Include *data on participants from 1990* or beyond.
Methods – Search Strategy

• Development of keywords
  ➢ Relevant to: developing nations, primary and secondary outcomes, RCT and QED evaluations

• Electronic searches of bibliographic databases

• Hand searches of relevant journals

• Citation tracking

• Contacting relevant authors and researchers

• Internet Searches and specialized holdings
Types of Interventions

- **Economic (n=26)**
  - Cash Transfers; Micro Finance; Labour Market; Tuition Relief etc.

- **Educational Programs and Practices (n=19)**
  - Remedial education, computers, flip charts, text books, and English language training technology and software

- **Health Care and Nutrition (n=14)**
  - Nutrition, treatment for asthma, malaria, vitamin A deficiency; school meals, etc

- **Building Schools and Infrastructure Improvements (n=7)**
  - Including new books; equipment; supplies, new roads, etc

- **Providing Information or Training (n=7)**
  - Livelihood skills, fertility control, parent training, community empowerment
Analysis

- Instrument designed to extract data from each study.
- Standardized mean differences effect sizes were computed for the first effect reported in each study, assuming random effects models.
- Main effects were analyzed for each outcome
- Meta-analysis was done to estimate overall mean effect size across studies, separately for different outcomes and across regions
‘Friendly Front End’ *(In Progress)*

- 10 new studies identified, screened for inclusion, assessed for quality, and analyzed as part of a modified update to the original review
- Update analysis of disaggregated/specific interventions
- Provide a clear set of policy issues and policy messages from the review
- In plain, accessible language
- With indications of what needs to be in place to achieve the positive outcomes that have been identified
Summary Results

• 73 included experiments and quasi-experiments in original review
• 10 new studies added in FFE
• Overall positive effect, on average, across all interventions

But:

• Aggregation of interventions - provides too gross a level of analysis (Type I and Type II errors are possible)
• Significant heterogeneity in effect sizes across all studies
Results

Average Effects by Broad Intervention Types

Largest Effects:
- New schools and other infrastructure interventions
- Health care and nutrition interventions

Smaller Effects:
- Educational programmes
- Information giving
Results

Figure 12. Average Effects Across Broad Intervention Types

<table>
<thead>
<tr>
<th>Group by Broad Intervention Type</th>
<th>Std diff in means</th>
<th>Statistics for each study</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic (n=26)</td>
<td>0.158</td>
<td>Variance: 0.001, Lower limit: 0.113, Upper limit: 0.204</td>
<td></td>
</tr>
<tr>
<td>Educational Practices/Programs (n=19)</td>
<td>0.043</td>
<td>Variance: 0.000, Lower limit: 0.015, Upper limit: 0.071</td>
<td></td>
</tr>
<tr>
<td>Health Care/Nutrition (n=14)</td>
<td>0.236</td>
<td>Variance: 0.004, Lower limit: 0.118, Upper limit: 0.354</td>
<td></td>
</tr>
<tr>
<td>New Schools/Infrastructure (n=7)</td>
<td>0.407</td>
<td>Variance: 0.002, Lower limit: 0.311, Upper limit: 0.503</td>
<td></td>
</tr>
<tr>
<td>Providing Information/Training (n=7)</td>
<td>-0.017</td>
<td>Variance: 0.001, Lower limit: -0.087, Upper limit: 0.054</td>
<td></td>
</tr>
</tbody>
</table>

-1.00 -0.50 0.00 0.50 1.00
Negative Positive
Results

Average Effects Across Regions

Largest Effects:
Studies that were conducted within:
• East Asia and the Pacific
• Europe or Central Asia

Smaller Effects:
Studies that were conducted within:
• Latin America and the Caribbean
• South Asia
• Sub-Saharan Africa
### Results – Across Regions

**Figure 15. Average Effects Across World Bank Classification of Developing Regions**

<table>
<thead>
<tr>
<th>Group by Region</th>
<th>Std diff in means</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia and the Pacific (n=4)</td>
<td>0.400</td>
<td>0.006</td>
<td>0.253</td>
<td>0.547</td>
</tr>
<tr>
<td>Europe and Central Asia (n=2)</td>
<td>0.619</td>
<td>0.044</td>
<td>0.209</td>
<td>1.030</td>
</tr>
<tr>
<td>Latin America and the Caribbean (n=25)</td>
<td>0.170</td>
<td>0.001</td>
<td>0.115</td>
<td>0.226</td>
</tr>
<tr>
<td>South Asia (n=23)</td>
<td>0.089</td>
<td>0.001</td>
<td>0.037</td>
<td>0.142</td>
</tr>
<tr>
<td>Sub-Saharan Africa (n=19)</td>
<td>0.096</td>
<td>0.001</td>
<td>0.038</td>
<td>0.153</td>
</tr>
</tbody>
</table>

**Statistics for each study**

- Std diff in means and 95% CI

- Negative
- Positive
Results – Effects by WB Economic Classification

Largest Average Effect:
• Lower Middle Income Countries (LMICs)

Smaller Average Effect:
• Lower Income Countries (LICs)

Smallest Average Effect:
• Upper Middle Income Countries (UMICs)
Results – Effects by WB Economic Classification

Figure 14. Average Effects Across World Bank Classification of Economies

<table>
<thead>
<tr>
<th>Group by</th>
<th>Std diff in means</th>
<th>Statistics for each study</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Bank Classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIC</td>
<td>0.124</td>
<td>0.001 0.080 0.168</td>
<td>-1.00 -0.50 0.00 0.50 1.00</td>
</tr>
<tr>
<td>LMIC</td>
<td>0.163</td>
<td>0.001 0.105 0.221</td>
<td></td>
</tr>
<tr>
<td>UMIC</td>
<td>0.115</td>
<td>0.002 0.035 0.195</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity statistics indicate that this is not a significant moderator (Q=1.39, df=2, p=.49).
# Results – Effects on Primary and Secondary Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Standardized Mean Effect (d)</th>
<th>BESD (Percentage Improvement in Treatment Over Control)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRIMARY:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolment</td>
<td>.18</td>
<td>9%</td>
</tr>
<tr>
<td>Attendance</td>
<td>.15</td>
<td>8%</td>
</tr>
<tr>
<td>Dropout</td>
<td>.05</td>
<td>3%</td>
</tr>
<tr>
<td>Progression</td>
<td>.13</td>
<td>7%</td>
</tr>
<tr>
<td><strong>SECONDARY:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>.16</td>
<td>8%</td>
</tr>
<tr>
<td>Language</td>
<td>.18</td>
<td>9%</td>
</tr>
<tr>
<td>Global Test Scores</td>
<td>.06</td>
<td>3%</td>
</tr>
<tr>
<td>Other Achievement</td>
<td>.05</td>
<td>3%</td>
</tr>
</tbody>
</table>

BESD = Binomial Effect Size Display (Rosenthal and Rubin, 1982)
Effects Across Evaluation Designs

Figure 16. Average Effects for Different Evaluation Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Std diff in means</th>
<th>Statistics for each study</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>QED (n=21)</td>
<td>0.141</td>
<td>Variance 0.001, Lower limit 0.080, Upper limit 0.201</td>
<td>-1.00 -0.50 0.00 0.50 1.00</td>
</tr>
<tr>
<td>RCT (n=52)</td>
<td>0.131</td>
<td>Variance 0.000, Lower limit 0.094, Upper limit 0.169</td>
<td>Negative Positive</td>
</tr>
</tbody>
</table>

- There was very little difference in these average effects ($d=.13$ for RCTs, $d=.14$ for QEDs).
Implications for Policy and Practice

• Interventions that address school enrollment, attendance, progression and dropouts have, on average, positive effects
• There are also positive effects on learning outcomes
• Effectiveness of interventions is *context specific*, and policy and practice should be implemented accordingly
• This requires a *theory of change analysis* – what activities, mechanisms, people, resources, and outputs are required
• And *disaggregated analysis* of specific interventions, countries and contexts
• Using qualitative and quantitative methods
• And cost-benefit/cost-effectiveness data and analysis
Thank you
Philip Davies
Email: pdavies@3ieimpact.org
+44 (0)207 958 8350

Visit www.3ieimpact.org