Examining the Experimental Designs and Statistical Power of Group Randomized Trials Funded by the Institute of Education Sciences

Jessaca K. Spybrook
A Presentation for the Evaluation Café at Western Michigan University
February 19, 2008
Background

- Evidence-based education
- Randomized trials
- Group randomized trials / Cluster randomized trials
Background

- Institute of Education Sciences (IES)
  - National Center for Education Research (NCER)
  - National Center for Education Evaluation and Regional Assistance (NCEE)

- Produce research that provides reliable evidence on which to base education policy and practice
Background

NCER

- Goal 3 Projects – Efficacy and Replication
  - Test effectiveness of intervention under specific conditions
  - ~ $250,000 - $700,000 per year

- Goal 4 Projects – Effectiveness Evaluations
  - Test effectiveness of intervention under more typical conditions
  - Up to $1.2 million per year
Background

- NCEE
  - Conduct rigorous evaluations of federal programs
  - Contracts not grants
  - At least $1 million per year
Background

- Group randomized trial ≠ Reliable, scientific evidence
- Strong design
- Large enough sample size to conclusively determine whether or not an intervention can improve student outcomes by a specified margin (adequate power)
- Power of 0.80 is usually considered acceptable in social sciences
Background - Terms

- **Minimum detectable effect size (MDES)** – Smallest effect size that can be detected with power $= 0.80$
  - Sample size at all levels
  - Intra-class correlation
  - Covariate-outcome correlation
  - Presence and strength of blocking variable
Central Goal of this Study

- Examine the designs and power analyses for the group randomized trials funded by the National Center for Education Research (NCER) and the National Center for Education Evaluation and Regional Assistance (NCEE)
Key Questions

1. What designs do these studies use?
Key Questions

2. Under plausible assumptions about intra-class correlations, covariate-outcome correlations, and explanatory effects of blocking, what are the minimum detectable effect sizes’s (MDES) of the studies in the sample?
Key Questions

3. What is the relationship between the MDES stated in the proposal and the MDES under plausible assumptions regarding the design parameters? To the extent that there are discrepancies between the two values, what are the possible sources of the inconsistencies?

- Is there a power analysis? Is it documented? Does it correspond to the study description?
- Are the intra-class correlations documented? If so, what are the estimated values?
- Are covariates included in the power analysis? If so, are the covariate-outcome correlations documented? If so, what are the values?
- Is blocking included in the description of the study? If so, is blocking included in the power analysis and are the explanatory effects of blocking documented? Is the treatment of the blocks (ie. fixed or random) stated, and if so, is it justified?
Sample

Pool of Studies

55 Potential NCER Studies
- 40 Received from direct contact with Principal Investigators
- 15 Sent request via FOIA and still waiting
  - 33 Meet criteria

13 Potential NCEE Studies
- 9 Received from NCEE directly
  - 6 Meet criteria
- 3 Received from direct contact with Principal Investigators
  - 3 Meet criteria
- 1 Sent request to Principal Investigator and still waiting
Sample

<table>
<thead>
<tr>
<th>National Center for Education Research</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal 3 Study</td>
<td>25</td>
</tr>
<tr>
<td>Goal 4 Study</td>
<td>8</td>
</tr>
<tr>
<td>National Center for Education Evaluation and Regional Assistance</td>
<td>9</td>
</tr>
</tbody>
</table>
Methods

- Classify the study design
- Determine plausible values for design parameters – intra-class correlations, covariate-outcome correlations, explanatory power of blocking
- Calculate the recomputed MDES
- Compare recomputed MDES to stated MDES
## Results – Experimental Designs

<table>
<thead>
<tr>
<th></th>
<th>Two-Level Cluster Randomized Trial</th>
<th>Three-level Cluster Randomized Trial</th>
<th>Three-level Multi-site cluster randomized trial(^a)</th>
<th>Four-Level Multi-site cluster randomized trial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Levels</strong></td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Level of Randomization</strong></td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Blocking?</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Number of Studies</strong></td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td><strong>Example of Nesting</strong></td>
<td>Students, Schools</td>
<td>Students, Classrooms, Schools</td>
<td>Students, Classrooms, Schools</td>
<td>Students, Classroom, Schools, Districts</td>
</tr>
</tbody>
</table>

\(^a\) Indicates a multi-site cluster randomized trial.
## Results – Experimental Design

<table>
<thead>
<tr>
<th>Experimental Design</th>
<th>Number of NCER Proposals</th>
<th>Number of NCEE Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-Level Cluster Randomized Trial</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Three-Level Cluster Randomized Trial</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Three-Level Multi-site cluster randomized trial</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Four-Level Multi-site cluster randomized trial</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>
Results - The Recomputed MDES

- Plausible values for ICCs
  - Bloom et al., 1999
  - Schochet, 2005
  - Hedges & Hedberg, 2007
  - Bloom, Richburg-Hayes, & Black, 2007
  - Murray & Blitstein, 2003
Results – The Recomputed MDES

- Plausible values for covariate-correlations
  - Bloom, Richburg-Hayes, & Black, 2007

- Plausible values for variance explained by blocking
  - Hedges & Hedberg, 2007
Results – Recomputed and Stated MDES

Solid Lines=Recomputed Effect Size
Dotted Lines=Stated Effect Size
Results

- Studies 1-24, MDES ranges from 0.40-0.90
  - NCER studies funded in 2002, 2003, 2004
  - Less likely to use a covariate

- Studies 26-J, MDES ranges from 0.18-0.40
  - NCER studies funded in 2005, 2006
  - NCEE studies
  - More likely to use a covariate
Results - NCEE

Solid Lines=Recomputed Effect Size
Dotted Lines=Stated Effect Size
Results - NCEE

- Recomputed MDES ranges from 0.10 – 0.40
- Majority of recomputed and stated MDES are in the same range
Results - NCER

Solid Lines=Recomputed Effect Size
Dotted Lines=Stated Effect Size

NCER Goal 3 Study
Results - NCER

Solid Lines=Recomputed Effect Size
Dotted Lines=Stated Effect Size
Results - NCER

- Similar for goal 3 and 4 studies
- Recomputed MDES ranges from 0.18 – 1.70
- Approximately half of the studies have recomputed and stated MDES in the same range
## Results – Relationship between stated and expected MDES

<table>
<thead>
<tr>
<th></th>
<th>Number of NCER Proposals</th>
<th>Number of NCEE Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDES within the same range</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Stated MDES &lt; Expected MDES</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Expected MDES &lt; Stated MDES</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

The 6 NCER studies without a power analysis are not included.
## Results – Details of Power Analyses

<table>
<thead>
<tr>
<th>Number of NCER Proposals</th>
<th>Number of NCEE Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same (n=14)</td>
<td>Stated&lt;Recomputed (n=12)</td>
</tr>
<tr>
<td>Simple statement of power with/without brief citation</td>
<td>6</td>
</tr>
<tr>
<td>Detailed power analysis with software or documented calculations</td>
<td>8</td>
</tr>
<tr>
<td>Optimal Design</td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
</tr>
</tbody>
</table>
## Results – Details of Power Analyses

<table>
<thead>
<tr>
<th></th>
<th>Number of NCER Proposals</th>
<th>Number of NCEE Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same (n=15)</td>
<td>Stated&lt;Recomputed</td>
</tr>
<tr>
<td>ICC estimate not included in proposal</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>ICC estimate included in proposal</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td><strong>Academic ICCs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within 0.10 to 0.20</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Not within 0.10 to 0.20</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Social or health ICCs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within 0.01 to 0.05</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Not within 0.01 to 0.05</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
## Results – Details of Power Analyses

<table>
<thead>
<tr>
<th>No covariate</th>
<th>Number of NCER Proposals</th>
<th>Number of NCEE Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same (n=15)</td>
<td>Stated&lt;Recomputed (n=11)</td>
</tr>
<tr>
<td>No covariate</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Covariate mentioned not documented</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Covariate documented</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>0.01-0.30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.31-0.50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.51-0.70</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.71-0.99</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### Results – Details of Power Analyses

<table>
<thead>
<tr>
<th></th>
<th>Number of NCER Proposals</th>
<th>Number of NCEE Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same (n=14)</td>
<td>Stated&lt;Recomputed (n=7)</td>
</tr>
<tr>
<td>Blocking included in the description</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Blocking included in the power analysis</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Include explanatory power of blocking</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Explicitly treat blocks as fixed effects</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Explicitly treat blocks as random effects</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Specify the effect size variability</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Conclusions

- Blocked designs are most common
  - Good for precision

- NCEE studies tend to have smaller MDES
  - Differences in funding
  - Differences in methodological guidelines
Conclusions

- NCEE studies tend to be more accurate
  - Training

- Growth is evident in accuracy and precision of NCER studies
  - More precise over time (use of covariates, blocked designs)
  - More accurate over time
Limitations

- Study proposals as data
- Use of original funded proposal