The Challenge of Authenticity in Scale-Up Effectiveness Trials

John F. Pane
Beth Ann Griffin
Daniel F. McCaffrey
Rita Karam

September 2011

This presentation has not been formally edited or peer reviewed and does not necessarily reflect the opinions of RAND research clients or sponsors
Features of a Scale-Up Effectiveness Trial

• Purpose: to determine whether an intervention with strong prior evidence of efficacy is effective at a broad scale
  – Authentic implementation
    • Under conditions that would be typical if a school district or other education delivery setting were to implement them (e.g. routine practice, without special support from the developer or research team)
    • As if the schools had purchased and implemented the intervention on their own without any involvement in a research study
  – In a diverse sample of schools, classrooms, or students to ensure appropriate generalizability

(source: IES Request for Applications, Education Research Grants, FY11, p. 72)
Authentic Implementation May Be the Greater Challenge

• Addressing generalizability is straightforward
  – Recruit from a sufficient variety of sites
  – Limited only by finding schools willing to participate, cost

• Some decisions about authenticity are also straightforward
  – For example, randomization of schools is likely to be more authentic than randomization of students within schools

• However, in many studies there are additional tensions between the demands of experimental methods and authenticity
  – In essence, the challenge is retaining control when authentic implementation requires schools to retain some discretion over student coursetaking
We Explore this Challenge in the Context of an Effectiveness Study of an Algebra Curriculum

• More than 24,000 students in 147 schools in 52 school districts in 7 states
  – Sites: Houston, TX; Bridgeport, CT; Trenton, NJ; Mobile, AL; suburban districts near Detroit; rural districts in Kentucky; districts throughout Louisiana
  – Includes urban, suburban, and rural schools, and some parochial schools (Catholic Diocese)

• Separate studies of middle schools and high schools
  – Population of students taking algebra in middle school is higher-achieving than students taking algebra in high school

• Each school participated for two years
Experimental Design

• School based randomization

• Blocking design where, within each site, schools were grouped into similar pairs and each pair was randomized to treatment or control groups
  – Variables used for pairing included: school size, racial composition, percent FRL eligible, percent ELL, three years of prior proficiency rates on state tests, and schools’ characterization of participating students

• Randomized experimental groups did not differ significantly on these school-level measures
## Participating Students by State and Grade

<table>
<thead>
<tr>
<th>State</th>
<th>Grade</th>
<th>Middle School Study</th>
<th>High School Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>6 7 8</td>
<td>9 10 11 12</td>
</tr>
<tr>
<td>CT</td>
<td></td>
<td>102</td>
<td>953 237 63 12</td>
</tr>
<tr>
<td>KY</td>
<td>3</td>
<td>49 1,352</td>
<td>4,350 282 37 14</td>
</tr>
<tr>
<td>LA</td>
<td></td>
<td>945</td>
<td>2,824 219 47 27</td>
</tr>
<tr>
<td>MI</td>
<td></td>
<td>77</td>
<td>2,799 318 112 66</td>
</tr>
<tr>
<td>NJ</td>
<td></td>
<td>36</td>
<td>1,094 86 18 19</td>
</tr>
<tr>
<td>TX</td>
<td>1</td>
<td>3,286</td>
<td>3,499 159 50 80</td>
</tr>
<tr>
<td>AL</td>
<td></td>
<td>871</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>50 6,669</td>
<td>15,519 1,301 327 218</td>
</tr>
</tbody>
</table>
Schools Typically Do Not Have a Well-Defined Population of Students Taking Algebra

- Majority of students take algebra in 9th grade
  - More advanced students may enroll in 7th or 8th grade
  - Lower achieving students may enroll in 10th grade or later
  - Some schools spread the curriculum over two years

- The definitions of these groups is not well specified and subject to the discretion of teachers, principals, or counselors

- Schools may be changing algebra enrollment due to pressures in some states to have most students take algebra by 8th grade

- Adoption of a new curriculum could be viewed as an opportunity to change algebra enrollment patterns
We Considered a Variety of Options for How to Control the Sample in the Study

<table>
<thead>
<tr>
<th>Control Mechanism</th>
<th>Level of Control</th>
<th>Authenticity</th>
<th>Feasibility Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify a grade in which all students must take algebra</td>
<td>High</td>
<td>Poor</td>
<td>Few schools would agree to participate</td>
</tr>
<tr>
<td>Require schools to specify precise student sample for both study years prior to randomization</td>
<td>High</td>
<td>Medium</td>
<td>Unacceptably high levels of attrition are likely</td>
</tr>
<tr>
<td>Administer algebra pre-post tests to all students in school</td>
<td>High</td>
<td>Excellent</td>
<td>High cost; schools might object to disruption of testing</td>
</tr>
<tr>
<td>Require schools to specify schema for student sample prior to randomization</td>
<td>Moderate</td>
<td>Good</td>
<td>Ability to monitor or enforce is limited</td>
</tr>
</tbody>
</table>
We Chose the Schema-Based Approach

• At formal enrollment in the study (prior to randomization), principals filled out this form

Designation of Study Participants
Please select one of the following rules to determine which students, classes, and teachers in your school will participate in the study during the 2007-08 and 2008-09 academic years. This rule will apply in both years. Any other algebra classes in your school will not participate in the study.

☐ All algebra classes.

☐ All algebra classes except ____________________________
(specify gifted, honors, remedial, etc.)

☐ All algebra classes taught by ____________________________
(specify teachers)
The students in these classes are (mark all that apply):
☐ Low performing
☐ Average
☐ High performing

Please estimate how many classes and students you have designated to participate:

Estimated number of participating classes: __________

Estimated total number of participating students: ______________
Potential Advantages of Schema-Based Approach

• Enables schools to implement the curriculum as they would have in the absence of the study
  – Good authenticity

• Provides some control over schools making changes to the study sample post-randomization
  – Important for internal validity of the experiment

• Enables us to account for variation in student participation in the pairing process
  – Can improve power of the study
Our Ability to Enforce This Was Limited

• We emphasized that schools must remain faithful to their designations for the two years of participation

• We periodically asked principals or department heads to confirm whether their designations were accurate
  – Some deviations were reported but overall we judged them to be unimportant

• The biggest problem is that even verbatim compliance with the specification allows for flexibility that is difficult to detect
  – For example, if a school specified “all algebra students” will be in the study, they might have made post-randomization decisions to increase or decrease the number of students taking algebra
We Assessed Balance on the Study-Administered Pretest

• Overall, the treatment group scored substantially lower than the control group on the pretest (ES=0.24, p<0.001)
  – Larger difference in middle schools than in high schools (0.35 vs. 0.15)
  – Present in both first and second years of participation (0.28 vs. 0.20)
  – Consistently true across all seven states, though not always statistically significant

• The What Works Clearinghouse threshold for meeting evidence standards is ES ≤ 0.25
Pretest distributions by group

Standardized pretest raw score

Density

-3 -2 -1 0 1 2 3

0.0 0.1 0.2 0.3 0.4

-3 -2 -1 0 1 2 3

Density

0.0 0.1 0.2 0.3 0.4

-3 -2 -1 0 1 2 3

Density

0.0 0.1 0.2 0.3 0.4

-3 -2 -1 0 1 2 3

-3 -2 -1 0 1 2 3

-3 -2 -1 0 1 2 3

0.0 0.1 0.2 0.3 0.4
What Might Have Caused Systematic Shifts in the Samples after Randomization?

- Pretest administered several weeks into the school year may have picked up an early treatment effect
  - Group differences also present on prior-year state test scores

- Treatment schools may have been enthusiastic about implementing a new curriculum and decided to increase enrollment of lower-achieving students in algebra
  - Treatment group has 14% fewer students than control group

- As administrators and teachers in treatment schools learned more about the new curriculum, they may have decided it would not be appropriate for higher-achieving students
  - Consistent with observing fewer students in treatment group
How We Plan to Handle this in Analysis

- The balance we hoped to achieve through randomization is called into question
  - Simple treatment versus control group analysis may produce a biased estimate if baseline covariates do not fully adjust for preexisting differences between the groups
- Will also run propensity score models that attempt to adjust for group imbalance
  - Will attempt to balance pretest scores and other covariates in treatment and control groups
  - Will attempt to balance groups within each matched pair
- Effects estimated by these two approaches can then be compared
## Design Option Three May Have Been Superior

<table>
<thead>
<tr>
<th>Control Mechanism</th>
<th>Level of Control</th>
<th>Authenticity</th>
<th>Feasibility Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify a grade in which all students must take algebra</td>
<td>High</td>
<td>Poor</td>
<td>Few schools would agree to participate</td>
</tr>
<tr>
<td>Require schools to specify precise student sample for both study years prior to randomization</td>
<td>High</td>
<td>Medium</td>
<td>Unacceptably high levels of attrition are likely</td>
</tr>
<tr>
<td>Administer algebra pre-post tests to all students in school</td>
<td>High</td>
<td>Excellent</td>
<td>High cost; schools might object to disruption of testing</td>
</tr>
<tr>
<td>Require schools to specify <em>schema</em> for student sample prior to randomization</td>
<td>Moderate</td>
<td>Good</td>
<td>Ability to monitor or enforce is limited</td>
</tr>
</tbody>
</table>
Implications for Designs of Future Effectiveness Studies

• For grades or subject areas where there is substantial discretion in whether students take a course or in what grade they take it
  – It may be necessary to collect outcomes for all students in the school, whether or not they take the course

• Such a strategy can maximize authenticity while retaining experimental validity, but
  – Can make the study substantially more costly
  – Might make schools less receptive to the study
  – There may be implications for study power