



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

- **School based randomization to new curriculum or business as usual**
 - **Blocking: schools were grouped into similar pairs and each pair was randomized to treatment or control groups**
 - **Randomized experimental groups did not differ significantly on available school-level measures**
- **Problem: schools typically do not have a well-defined population of students taking algebra**
 - **Most students take algebra in 9th grade, advanced students in 7th or 8th grade, lower-achieving students in 10th grade or later**
 - **Decisions are discretionary**
 - **Some states are pushing for algebra by 8th grade**
 - **Change in curriculum could spur changes in algebra enrollment**

We Considered a Variety of Options for How to Control the Sample in the Study

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

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 $ES \leq 0.25$
- Will conduct supplemental analyses to address imbalance

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**

Implications for Designs of Future Effectiveness Studies

- **For grades or subject areas where there is substantial variation 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**

