Research Strategies and Designs for Evaluating AMSTI and Similar Programs

Denis Newman

Empirical Education Inc.

Presentation at NCSL meeting on STEM

Huntsville, AL

January 10, 2009

REL SOUTHEAST
Regional Educational Laboratory

Who is Empirical Education

- Founded 2003 in Palo Alto CA. Conducting research in all parts of the US
- Focused on evaluating K-12 instructional and professional development programs
- Currently working under contract with REL-Southeast to conduct the evaluation of AMSTI's impact
- Began work in Alabama in January 2006 under funding from a separate IES grant

Goals for this Presentation

- Will not be reporting the results of the AMSTI experimental evaluation
- Focus is on what we learn from conducting a large scale study on an important STEM initiative
  - What can be found out from experimental evaluations
  - How states can conduct and use rigorous research
  - Implications for state data systems

Why Evaluate AMSTI

- Benefit to the State:
  - Provides feedback to improve the program
  - Measures the impact in relation to the State high stakes test in Science, Math, and Reading
  - Measures differential impacts on sub-populations
  - Uses a rigorous method that will provide a better measure of the impact than prior studies that use non-experimental methods
- Benefit to the Researchers
  - The “research site” is also the client
  - Provides an opportunity to work with an ongoing program
  - *And* to work with schools new to the program—required for randomized experiments

Basics of the AMSTI Experimental Evaluation

- Following 82 schools over three years
- Looking for impact of AMSTI on state-administered tests of
  - Science
  - Math
  - Reading
  - Also, surveys of teaching practices
- Empirical Education has now received the student data from 2007-2008 school year
  - Data analysis now in progress
- But first, some basics of experimental evaluation
  - Using real data (not from AMSTI)

REL SOUTHEAST
Regional Educational Laboratory

The Classic Elementary Science Experiment

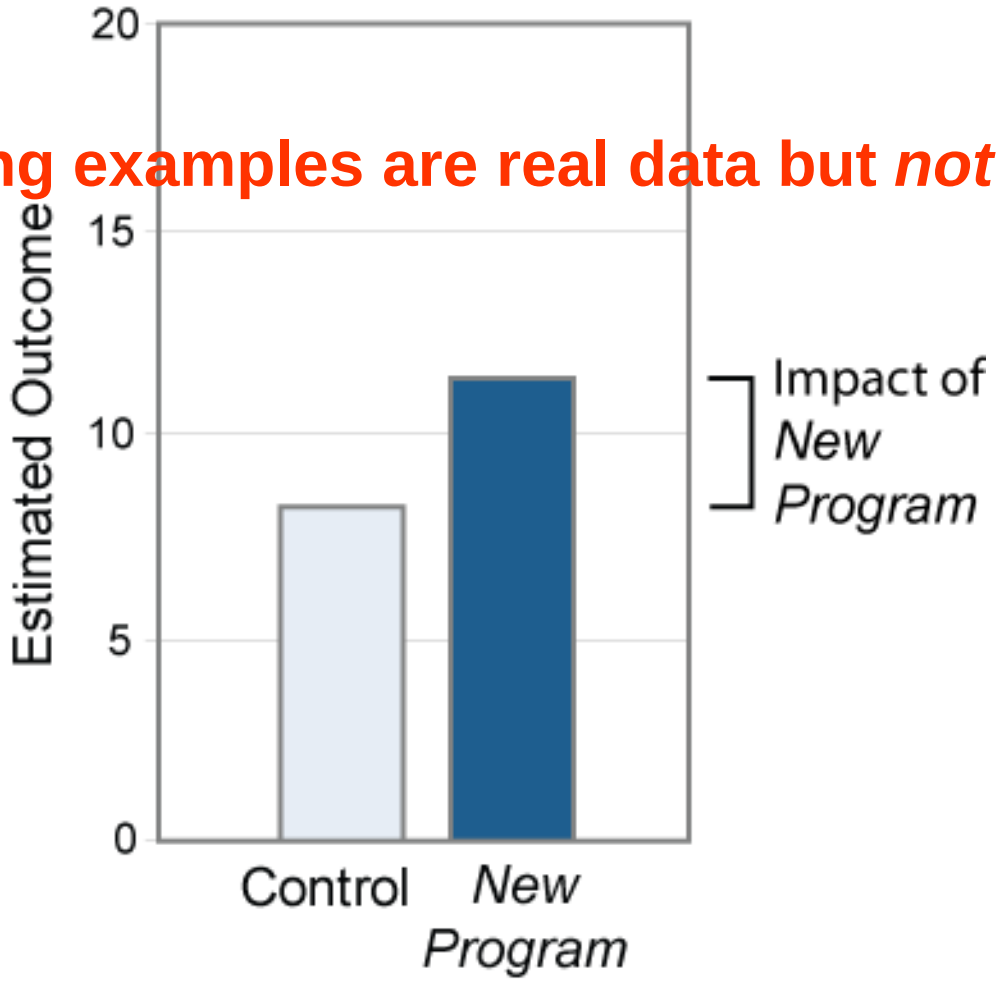
- Put one tomato plant in the closet, leave the other in the sun—see what happens!
- Compare a group that had a new program to one that didn't
- Estimate the difference between the two = *evidence of impact*
- Two must be equivalent (or very similar) to start with

Why We Use a Control Group

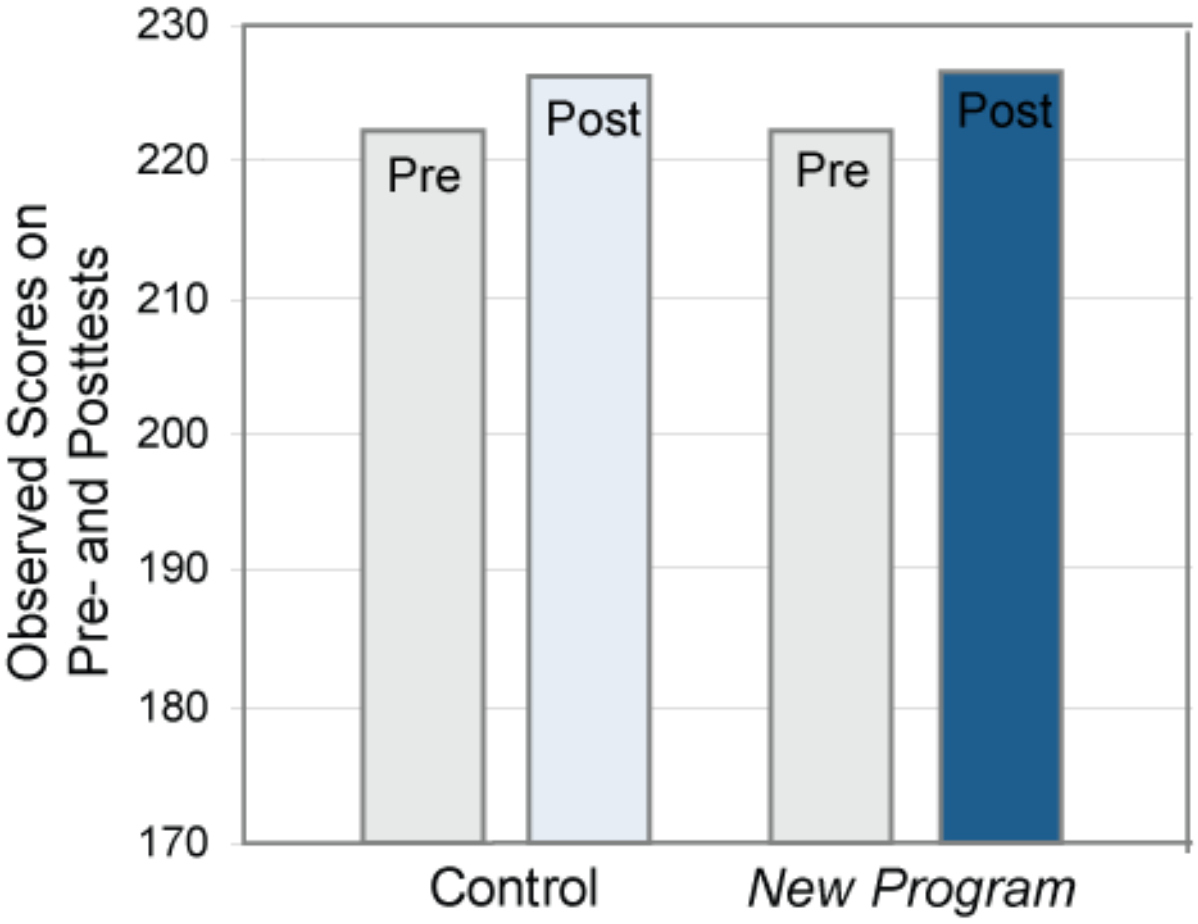
- Basic question: is the new program any better than what you already have?
- Control group
  - Represents what the program group would have looked like at the end of the experiment if it didn't get the program
  - In education this is almost always some existing program or "business as usual"
- Choosing the program and control groups with a lottery or coin toss is the best method to assure the groups are "interchangeable".

Impact: The Difference the New Program Makes

Note: following examples are real data but *not from AMSTI*

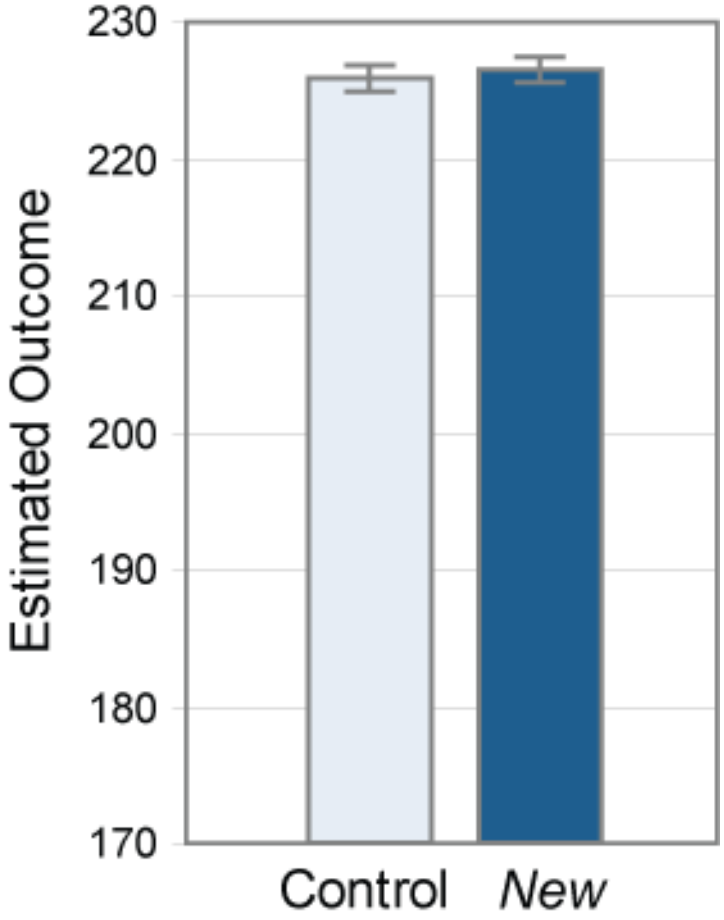


Growth from Pretest to Posttest

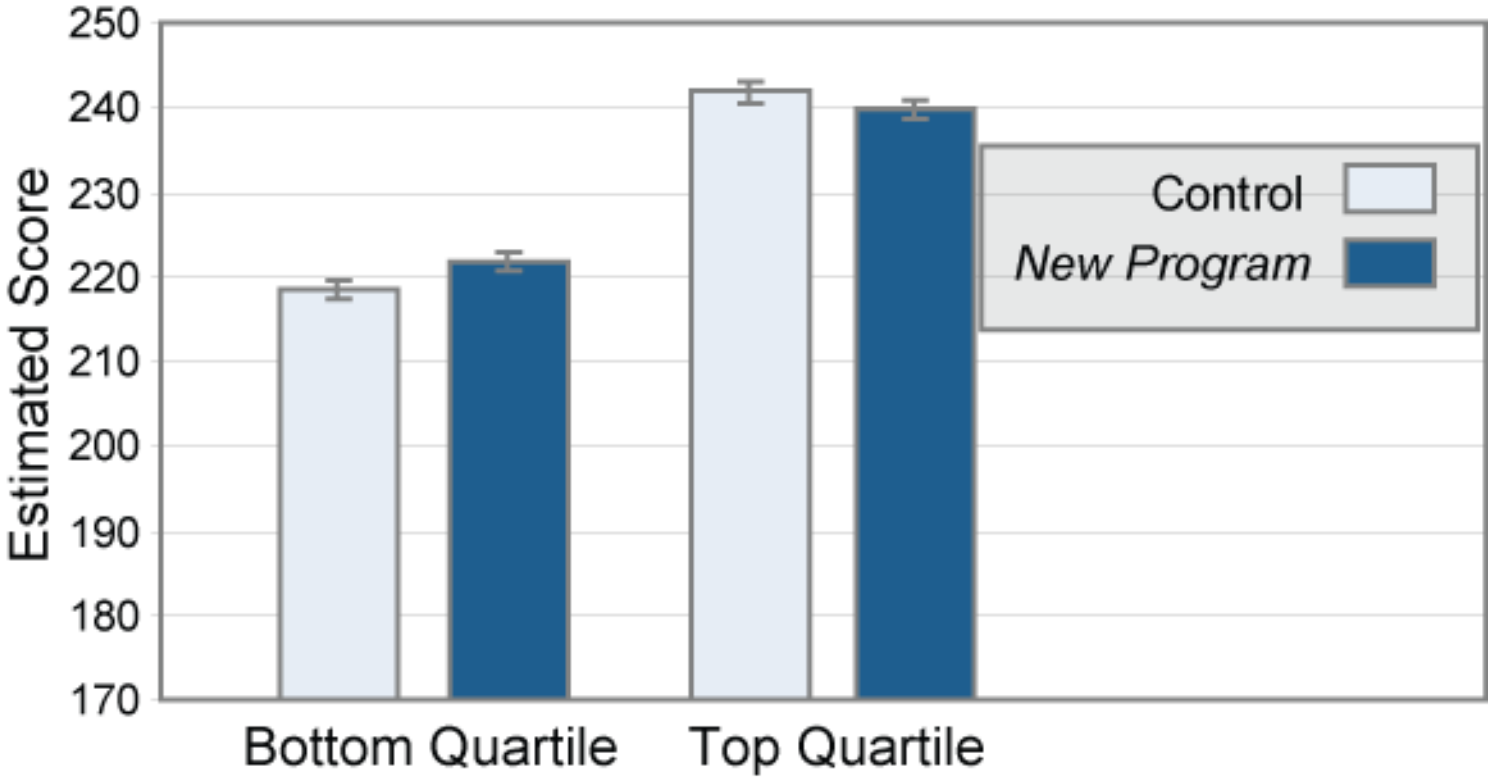


REL SOUTHEAST
Regional Educational Laboratory

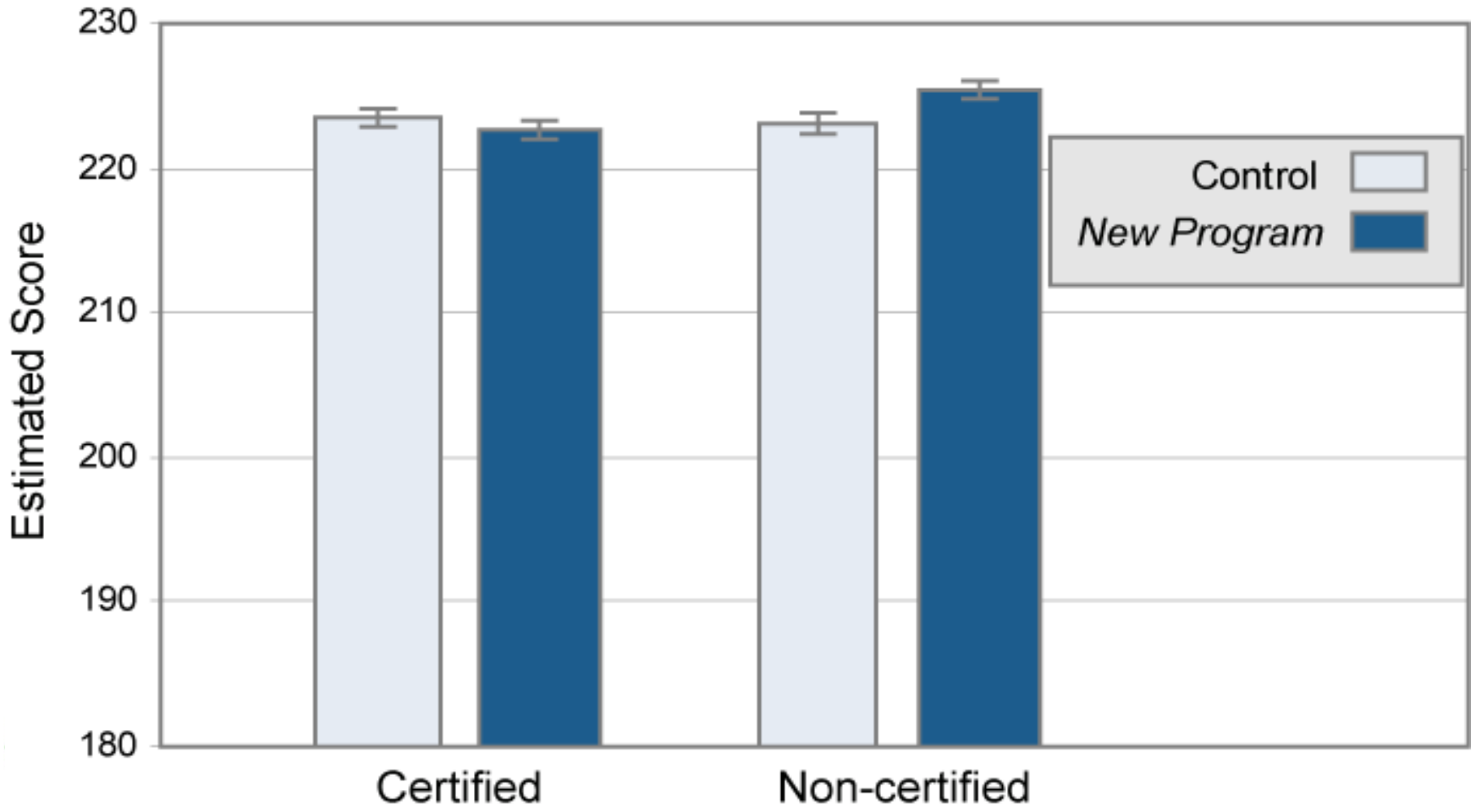
But No Difference Between the Groups



However…difference between bottom and top students



And, differences depending on teacher preparation



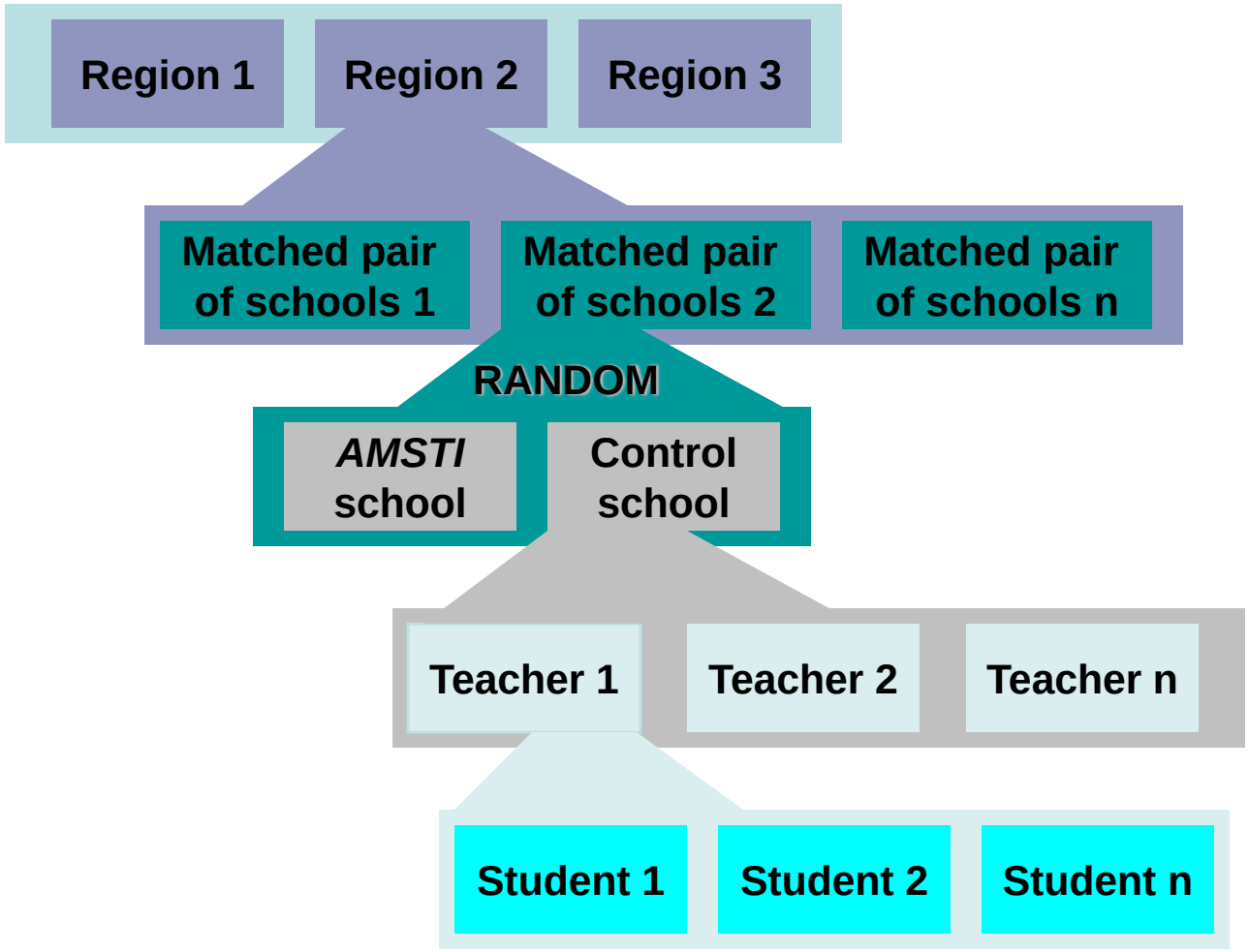
The AMSTI Experimental Evaluation

Taking advantage of

- Incremental scaling up
  - New regions came on board each year
  - For new regions, we could randomly assign schools to start AMSTI now or start in a year
- Greater demand than supply
  - More than 100 applicants for about 20 slots
  - Schools agreed to participate in research as part of the application

Random assignment

- Schools assigned to start AMSTI this year or next year
- Coin tossed between pairs of similar schools


Region 1
Region 2
Region 3
Matched pair of schools 1
Matched pair of schools 2
Matched pair of schools n
RANDOM
AMSTI school
Control school
Teacher 1
Teacher 2
Teacher n
Student 1
Student 2
Student n

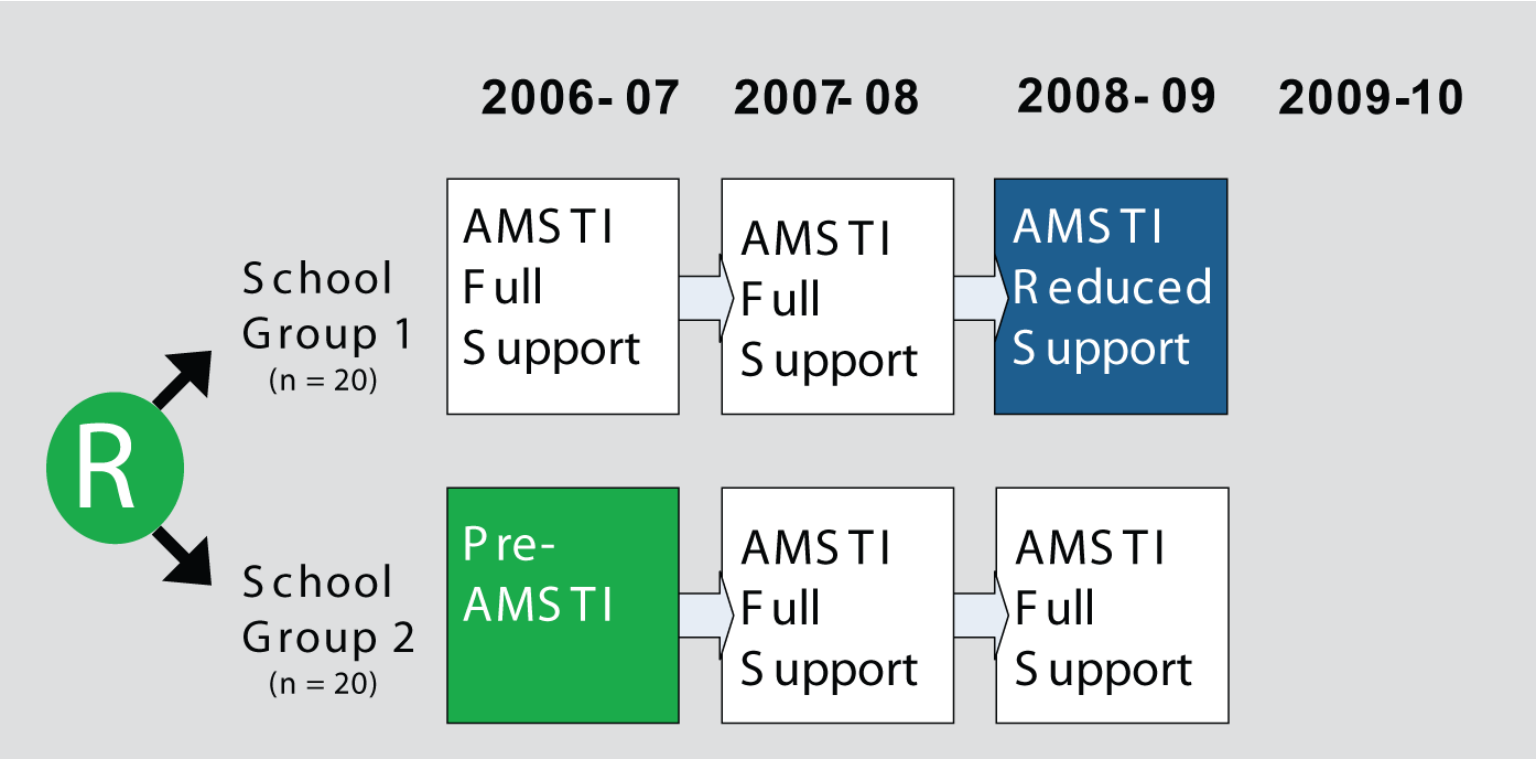
Mathematics Participant Counts by Schools, Teachers, Classes and Students for 06-07

	Randomized Schools	Teachers	Classes	Students
Control	20	103	232	5408
AMSTI	20	134	256	5862
Totals	**40**	**237**	**488**	**11270**

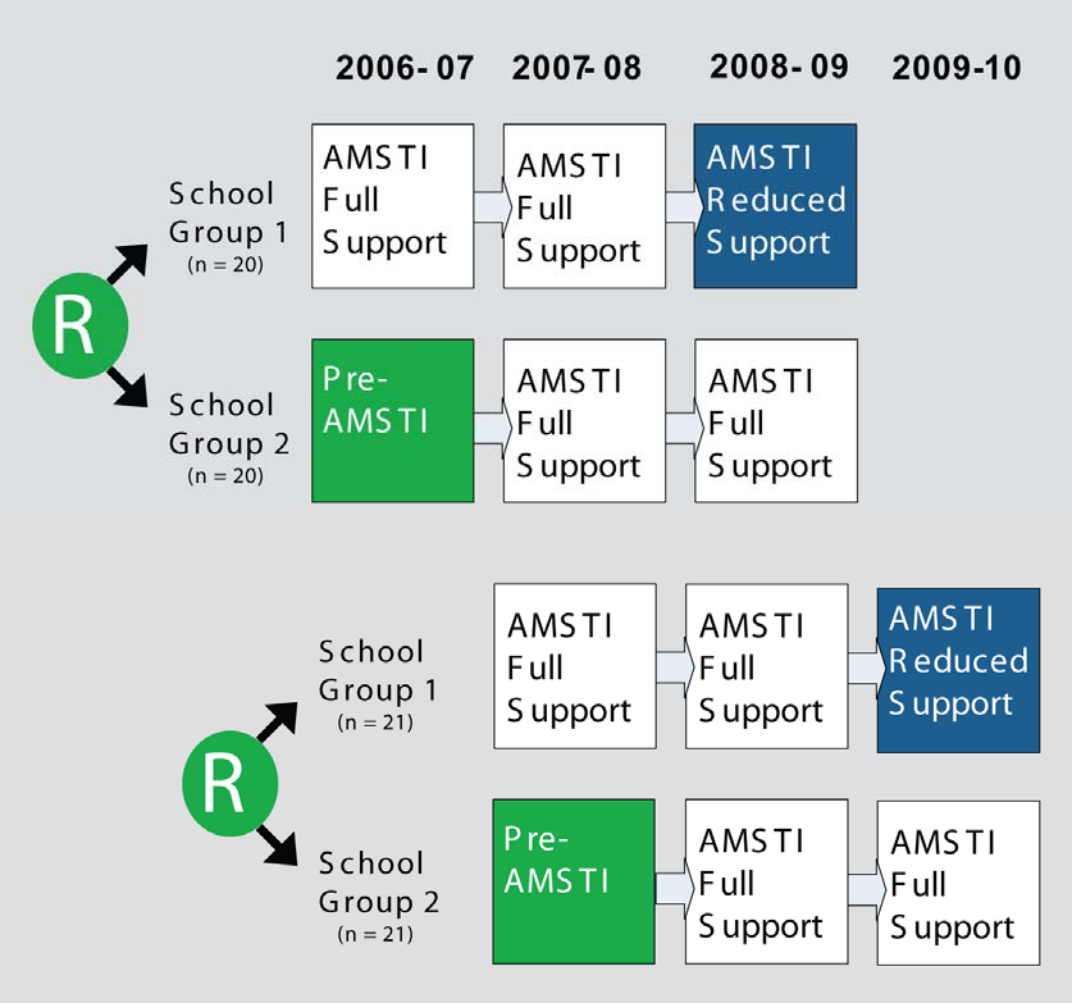
Mathematics Participant Counts by Grade Levels for 06-07

	Number of Students					
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
Control	749	685	1246	1317	1411	5408
AMSTI	1011	1414	1091	1170	1176	5862
Totals	**1760**	**2099**	**2337**	**2487**	**2587**	**11270**

The AMSTI Experimental Evaluation



The AMSTI Experimental Evaluation



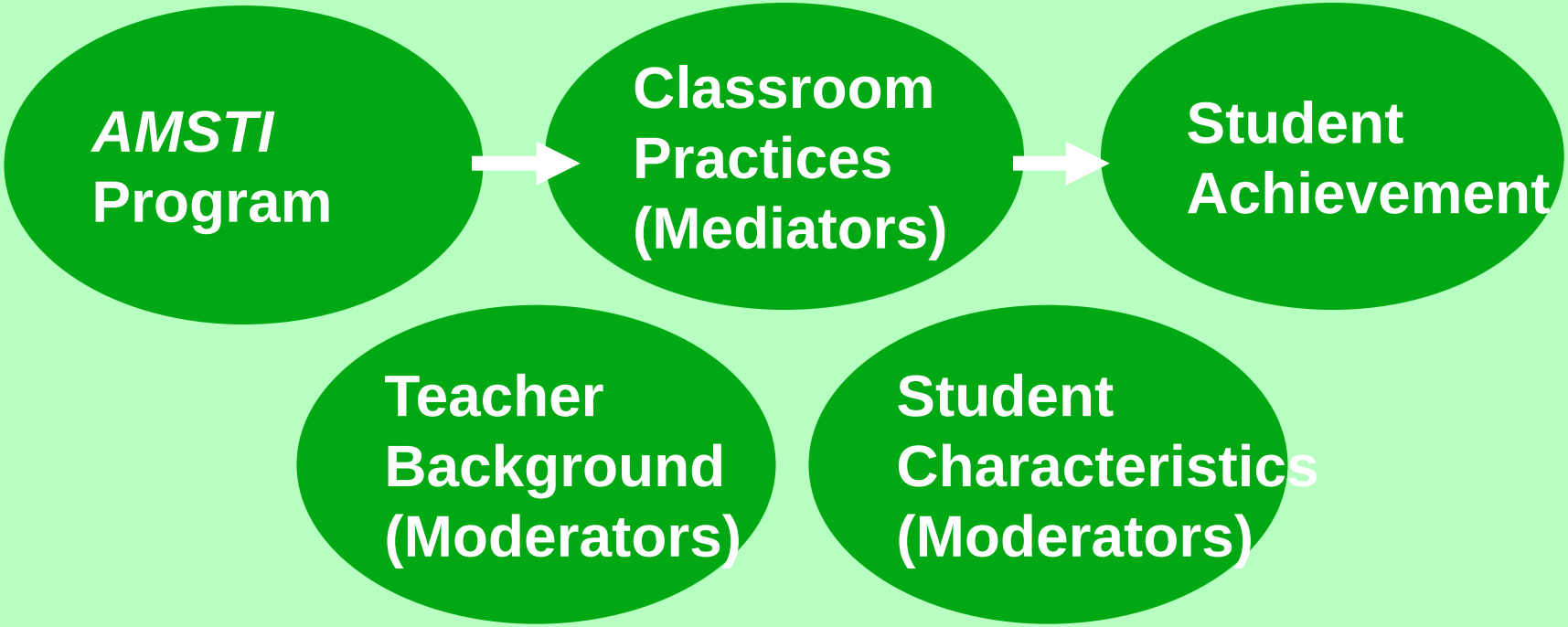
Theory Behind AMSTI Experiment

Professional development, materials, technologies, in-school supports

Inquiry processes, hands-on activities

Mathematics, Science & Reading

Theory Behind AMSTI Experiment



Levels of Analysis are Important

- We are looking at
  - Schools
  - Teachers
  - Students
- Student characteristics may appear to “moderate” the impact, but
  - We can look for organizational effects such as school or community characteristics
  - Role for theory—must avoid fishing for correlations
- Important to understand how students are clustered
  - Within a school or with the same teacher
  - Some statistical techniques have it built in

Pros and Cons of Randomized Experiments

Pros:

- Eliminates the most important sources of bias
- Sets up a collaboration between the researchers and both the program and the control schools

Cons:

- Requires planning prior to implementation
- Always starts with a new implementation—takes time to get the initial results

"Quasi-experimental" Alternatives for Evaluations of State and District Programs

- Select a comparison group from among very similar schools

Pros:

- Can look at a program already underway
- Can get information on impact faster (especially looking at multi-year impact in retrospect)

Cons:

- Very hard to avoid bias resulting from the selection of groups into the program
- Hard to get comparable information from people in the comparison group
- Real-time data collection often unavailable

Implications for State Data Systems

- Longitudinal systems
  - For following multi-year programs
  - Useful to have testing that uses a growth model
- Link between the student and teacher
  - If we want to understand
    - Impact on classroom practice
    - Association of impact with teacher preparation
  - School districts will have the most accurate class rosters
  - In Alabama and elsewhere—a two step approach: district then state
- Universal student identifier
- Balancing privacy and value of scientific research

State and Local School Systems as Scientific Investigators

- States, districts, teachers doing science—an excellent way to get the STEM message across
- The local control group is what is relevant to local policy decisions
- Local questions may be of particular importance:
  - A particular achievement gap
  - Impact on specific state tests
  - Remember: researchers have to choose a limited number of hypotheses so select ones of local interest
- Using the scaling up process as an opportunity for continuous feedback and improvement

Denis Newman
President
Empirical Education Inc.
dn@empiricaleducation.com
(650) 328-1734 x112
www.empiricaleducation.com