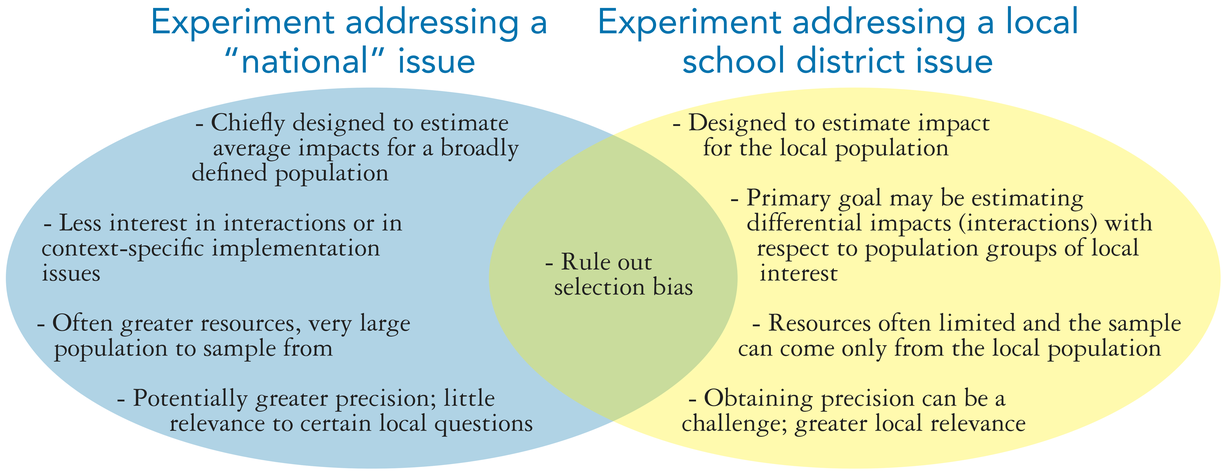
# Scaling Down: Low Cost Experiments to Address Questions of Local Relevance

Denis Newman and Andrew P. Jaciw
Empirical Education Inc.

## Questions.

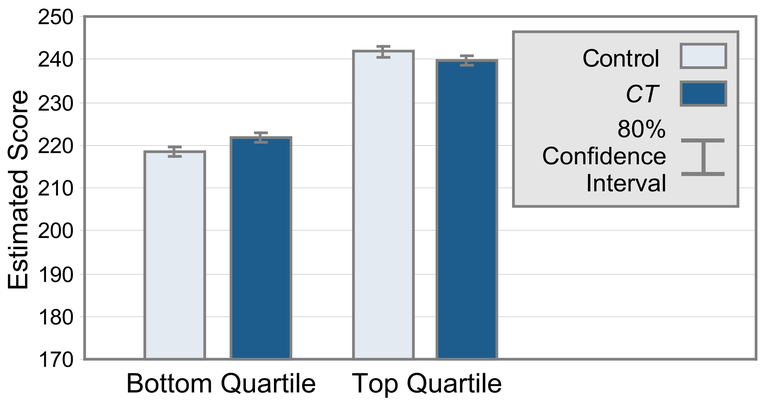
1. Can local experimentation become a routine part of how school districts inform their decisions?
2. Can experiments be inexpensive, informative, timely, and relevant to local issues?




## Example of work.

During the three years of the project, eight RCTs were initiated under a grant from U.S. Department of Education. For example, we worked with a school system in Hawaii to test the effectiveness of Carnegie Learning's *Cognitive Tutor* for pre-algebra during the 2006-2007 school year. Important characteristics of this study:

- Since there were only 12 teachers, we randomized classes within teachers to obtain 32 units and sufficient power.
- An important question for the schools was the gap between native Hawaiians and others. Native Hawaiians scored lower on the pretest, therefore measuring the differential impact for low and high scoring students allowed us to consider the relative effectiveness for Hawaiian students.
- While we did not detect an average impact, this interaction's *p* value was .02. The results also suggested that *CT* had more value for uncertified than for certified math teachers.




In addition to the eight US ED-funded RCTs, we conducted another 13 RCTs using commercial and other funding.

1. Middle School Social Science Resource (2004-2005)
2. Elementary Science Resource (2005-2006)
3. Computer-based Teacher Support System (2004-2006)
4. *Cognitive Tutor* for Algebra 1 (2005-2006)
5. Professional Development for Interactive Whiteboards (2005-2006)
6. *Cognitive Tutor* for Pre-algebra (2006-2007)
7. Middle School Math Tutoring (2006)
8. A state-wide Math and Science Initiative (2006-2010)

## Technical issues addressed during the project.

Low-cost and locally-relevant experiments pose specific challenges. Foremost is obtaining sufficient statistical power to detect average and differential effects. To increase power, we use matched pairs designs and randomize classes or teachers rather than schools (if we can assess that randomizing units at lower levels does not unduly disrupt normal school processes.) We also consider district officials' tolerances for drawing false-positive and false-negative conclusions as a basis for choosing settings for type-1 and type-2 error rates in the power analysis. Further, we consider an alternative to the Minimum Detectable Effect Size (such as a Minimum Required Effect Size) for determining sample size needs.

## Logistical/Organizational challenges.

How can the cost of local experiments be lower compared to national experiments?

- Eliminate cost of recruiting and reduce costs by working within a single jurisdiction.
- Develop "standard operating procedures" for conducting data collection, surveys, statistical analysis, and reporting.

How can results be timely?

- A major challenge is turning results around to be useful for a decision about the upcoming school year.
- Experimental evaluation and program implementation must be planned concurrently.

How can we assure the results are used?

- Most districts are not prepared to use locally generated evidence in decisions about programs.
- Our conclusion from the project is that local experimentation represents a reform innovation building on and going well beyond current data systems.

## Addressing technical challenges.

In a continuing effort to address technical challenges to conducting low-cost and locally-relevant experiments, we are planning the following activities:

- Inventory estimates of critical parameters for doing power analyses
- Investigate power for detecting differential effects (power for differential impacts among subgroups of students may be greater than for detecting average effects of comparable size)
- Investigate alternative randomization schemes (e.g., randomizing grade levels)
- Empirically investigate thresholds for obtaining benefits from using matched-pairs designs
- Investigate statistical power for detecting mediator effects and impacts under adequate implementation
- Determine the potential of value-added scores as supplementary covariates for increasing precision

## Next steps.

**Evidence-based decision-making is a *systemic reform* not just a decision procedure.**

- Develop a reform model for making educators in districts active participants in shaping their own research questions.
- Enhance professional development to move district leaders from data users to evidence producers. Help them build the internal capacity for producing usable evidence.
- Deploy technologies for supporting research capacities in districts and schools—see Empirical Education's MeasureResults poster!

The research was funded by a grant (#R305E040031) to Empirical Education Inc. from the U.S. Department of Education.