Lessons Learned from Two Large Scale RCTs: Similarities and Differences

Symposium Justification

Strong designs such as experiments facilitate causal inferences and establish rigor in educational research. Well designed and executed experiments have the potential to provide causal claims about the effects of school interventions on student achievement. However, even robust experimental designs do not always yield definitive answers about what works in schools. In fact, mixed findings are not uncommon in educational research. For instance, the effects that small classes or effective teachers have on student achievement have varied in the literature. In addition, work on charter schools, after-school programs, and teacher professional development have typically yielded mixed findings. Over the last decade there has been a thrust in education research to conduct large-scale experiments to evaluate the effects of school interventions. Yet, even rigorous evaluations of school programs may produce inconclusive findings.

The proposed symposium speaks directly to the theme of the conference, Capitalizing on Contradiction: Learning from Mixed Results, providing insights that will help the design and execution of future experiments. It consists of a collection of papers that discuss contradictory findings produced from two, independent large-scale experiments that were conducted in public schools in Indiana in two consecutive years (2009-2010 and 2010-2011). The experimental design was unique because it incorporated a replication study (i.e., the experiment in the second year 2010-2011) of the evaluation of diagnostic assessments tools. The experiments were cluster randomized designs, that is, schools were assigned randomly to a treatment (i.e., diagnostic assessment) or a control (i.e., business as usual) condition. Each experiment was large and included nearly 50 schools and thousands of students in grades k through 8. Schools were randomly selected from lists of schools that had volunteered to implement the intervention the following academic year.

The first paper describes the two experiments and juxtaposes the results of the first and second years of the study. Further, the paper discusses the findings and attempts to provide some insight into the differences between the findings. The second paper focuses on the inner workings of the two experiments, looking specifically at the control condition. Using surveys and logs of instructional practices, the findings show that both treatment and control schools used numerous assessment systems before the study began and that control schools continued to use a wide range of assessments during the study. The third paper examines a primary intermediate outcome, the degree to which treatment and control group teachers used differentiated instruction to assist student learning (in both experiments). Specifically, the third paper used teacher logs of instructional practice to construct a measure of differentiated instruction using IRT models. The comparisons between treatment and control schools indicated that teachers in control schools differentiated instruction significantly more than teachers in treatment schools. Finally, the fourth paper examines the external validity of the findings in the two experiments using recent methods that use propensity scores. Through this group of papers, the symposium offers a deeper look into the studies’ contradictory findings, providing insight that will help the research community best design and implement future studies.

Lessons Learned from Two Large Scale RCTs: Similarities and Differences