

Abstract Title Page
Not included in page count.

Title: Lessons Learned About Random Assignment Evaluation Implementation In the Classroom

Author(s): Savitha Moorthy, Ph.D., Raquel Sanchez, Ph.D. and Fannie Tseng, Ph.D.

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

The current climate of accountability has ushered in a focus on evidence-based research, especially to inform decisions about the deployment of large-scale interventions (National Research Council, 2002; Slavin, 2002). In this context, rigorous experimental designs -- RCTs -- that measure causation and clearly illustrate the relative effectiveness of different interventions, have taken center-stage in educational research. No Child Left Behind (US Congress, 2001), for example, strongly emphasizes “scientifically based research” defined as “rigorous, systematic, and objective procedures to gain valid knowledge”, and operationalized typically as experimental or quasi-experimental designs, preferably with random assignment. Cook and Payne (2002) believe that education research has been weak at establishing causal relationships; this weakness, they claim, is based on contemporary researchers’ objections to using experimental studies to understand such relationships. They maintain that critiques of RCTs do not provide a convincing rationale prohibiting their use in educational settings, as most of the issues highlighted above can be resolved. Therefore, despite reservations, the literature overwhelmingly recommends the use of randomized trials often in conjunction with equally rigorous qualitative research, to establish causality and identify the relative effects of interventions.

Gueron (2002) suggests that operational and political skills are as important to successful execution of random assignment as technical skills and rigorous methodology. A thoughtful design phase in an RCT, which attends to local stakeholders and incorporates their needs and concerns, can pave the way for a successful study (Raudenbush, 2005). To secure the buy-in and involvement of school administrators prior to mounting an RCT, for example, Kane et. al. (2008) offer the example of involving them in methods discussions during the early stages of school-based randomized trials. Gueron (2000; 2002) also counsels researchers to ensure that they are asking the right question, and meeting ethical and legal standards.

The implementation of RCTs in real world educational settings poses practical challenges. They are more expensive to implement and, for classroom-based interventions, the number of schools and classroom needed to ensure adequate statistical power is logistically challenging. Moreover, it is often difficult to secure buy-in from teachers and administrators, which has implications for the implementation fidelity and quality.

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

The objective of our paper is to provide some practical insight into the process of implementing large-scale randomized controlled trials (RCT) in educational settings and to provide general recommendations for researchers conducting education-based RCTs in the future. We draw our recommendations from our recent experiences implementing RCTs in classroom settings for four Department of Education-funded evaluations. Our paper highlights the complexity of, and the lessons learned from, implementing large-scale experimental evaluations in real world contexts.

Our findings are in concurrence with the previous literature on implementing RCTs (for example, Gueron (2002) and Kane et al. (2008)), which has stressed the importance of political and operational skills in the successful execution of RCTs. In this paper, we expand upon the previous literature by examining and describing how RCTs articulate in concrete educational settings, where the stakeholders include school districts, school administrators, teachers, students, and parents. Here, we highlight the specific issues that need to be taken into careful consideration while evaluating classroom-based interventions, where the intervention is not administered directly to student, but rather through the participating teacher. Finally, we highlight the lessons learned from our experience of implementing four RCTs, and discuss the strategies evaluators can employ to deal with and mitigate the complexities involved with implementing RCTs in classroom settings.

Setting: Four RCTs conducted in the states of Arizona, California, Florida, Texas and Illinois and the U.S. Territories of American Samoa and Commonwealth of the Northern Mariana Islands (CNMI).

- **Population / Participants / Subjects:** Participants in the four RCTs include classroom teachers and:
 - Elementary School English Language Learner (ELL) Students;
 - High School ELL Students;
 - Adult ELL Students;
 - Infants and Toddlers, ages 0-2.

Intervention / Program / Practice: Description of each of the four programs under evaluation using RCTs:

2. A professional development program for secondary English as a Second Language (ESL) and English Language Arts (ELA) teachers that is an integrated package of summer institutes, in-person one-on-one coaching sessions during the school year, and collaborative lesson/unit planning activities.
1. A professional development program for 4th and 5th grade teachers in Hawaii and other Pacific Islands to improve their pedagogical knowledge and classroom practices with regard to literacy instruction for English language learners.
3. A literacy instruction and training program designed to improve the reading and language skills of low-literate adult ESL students.
4. A subsidized training and technical assistance program for infant/toddler center, family child care, and licensed exempt providers that utilizes a relationship-based approach to infant/toddler care.

Research Design: Qualitative case study of four classroom-based RCT studies, using RCT implementation experiences to draw up a set of recommendations for conducting RCT studies and dealing with the challenges of implementing an RCT “on-the-ground”.

Data Collection and Analysis: Project director interviews and analysis of preliminary implementation study results.

Findings / Results:

During the past four years, we have implemented random assignment evaluations of school-based interventions in over 450 classrooms. In the process, we have encountered situations and events that have posed challenges to the evaluation. In this section, we will describe some of these challenges in order to illustrate the contextual factors that can make implementing a random assignment evaluation on the ground difficult and, when possible, provide some strategies to mitigate the potential negative consequences of these challenges on the evaluation.

1. Threats to the Integrity of Random Assignment

We have found that it can be a challenge to maintain the integrity of random assignment in school-based settings, where there is frequent turnover both of students and teachers. Moreover, depending on the study design, there is sometimes a long time-lag between the RA assignment and the implementation, and it is not always possible to limit movement of participants during this waiting period. In these cases, the integrity of random assignment is threatened by crossovers, where program students switched classrooms and became control group students, and/or contamination, where the intervention spread to control group sites by program-trained teachers who switched classrooms or schools. Because the movement of students and teachers within schools and districts is unavoidable (and, to some extent, inevitable), sample loss due to crossovers should be anticipated and factored into the design models of school-based studies. Contamination is a larger problem for which there is no simple solution, since whole schools and districts are affected. However, we believe that the potential threat of contamination illustrates the importance of conducting a thorough and rigorous implementation study, through which threats like contamination can be detected.

2. Recruiting and Obtaining Buy-In

Another major issue with implementing school-based random assignment evaluations is the large number of schools required to ensure adequate statistical power for the design. The average number of schools that were recruited in each of our classroom-based random assignment evaluations was 50 (?). For any organization, recruiting this many schools is a major undertaking that requires significant time and resources. Recruiting at the district-level is more efficient, because one can recruit multiple schools in a district with visits to district-level administrators, and in most cases, obtaining buy-in and permission from the district administrator is a necessary condition to doing the study. However, there are no shortcuts to site recruitment; once district-level buy-in is obtained, in order to ensure buy-in from the implementers and study participants, we have found that it is essential to communicate, and develop and maintain relationships, with stakeholders at the school-level as well as the district level. Because none of the necessary data collection activities can effectively take place if we have not established and maintained good working relationships with study participants, proceeding with the evaluation without obtaining buy-in from principals and teachers can result in difficulties during the data collection phase of the evaluation.

3. Dilution of Intervention Effectiveness Due to Lack of Teacher Buy-In

One common feature of the school-based interventions we have evaluated is that the intervention being tested is a professional development program administered at the teacher-level. Although, we evaluate the intervention by assessing their impact on student-level outcomes, the success of the intervention actually depends on two factors: (1) the effectiveness of the intervention and its implementation at the training-level (program-to-teacher level) and (2) effectiveness of the teacher's implementation of the training at the teacher-level (teacher-to-student level). If the implementation of the intervention breaks down at either of these levels, the effectiveness of the intervention will be compromised.

We have found that because implementation of the professional development intervention depends so crucially on the teacher and her/his ability to transfer the learned skills and training to her/his students, "on model" implementation is actually somewhat hard to achieve. Moreover, if the program group teachers are not motivated to participate in professional development training and implement the skills and knowledge learned in the training, the intervention will also be less likely to succeed. For example, one random assignment evaluation conducted in California has coincided with the State budget crisis which affected teachers in districts throughout the State. Not only were program group teachers were laid off, understandably, the morale of the program group teachers who remained was not high; in this case, teachers participating in focus groups stated that, given the state's fiscal situation, their participation in the study was not a high priority. In another case, we found that a lack of buy-in of teachers participating in the study due to disputes between teachers and school administrators compromised the quality of program implementation.

One practical implication of weak program implementation is that the study may end up being underpowered for the impact analyses – the expected effect size for the intervention might end up being lower than what was assumed during the design phase of the evaluation, and as a result, the study sample size may be too small to detect the lowered effect size.

While we have discussed above some steps evaluators can take to maximize the chances for buy-in at the district and school-levels, because effective program implementation depends so critically on the performance and motivation of the teacher, there are times when one cannot control how well the program is implemented. In our paper, we provide recommendations for how evaluators can prepare in advance for these types of scenarios.

4. Keeping Schools in the Study After They Have Abandoned the Intervention

While retaining schools and districts in the study has not been a significant problem for us, there have been a few cases where sites have decided not to continue with the treatment after the first year of the study. In these cases, the burden is on the evaluator to keep the school as a participant in the study and continue to collect data as part of the intent to treat model through the end of the study. In order to facilitate these discussions, we recommend that evaluation staff try to establish an independent identity from the program they are evaluating from the outset so that the evaluators can develop its relationship with the district beyond the immediate project. In our paper, we will provide recommendations for how evaluators can establish an independent identity while also maintaining a close and productive relationship with program developers.

5. Documenting Treatment Dosage

We have also encountered problems with documenting the dosage of the treatment. Given that this research design comes from a medical model for research, educators often do not understand nor value the importance of dosage data, and programs often do not keep very accurate or detailed records that would allow us to document dosage reliably. From these experiences, we have learned that we need to be very proactive about our plans for documenting dosage and explain those plans to the program developers up front. Because dosage is a crucial variable in the impact analysis, we have learned that we cannot afford to leave this data point to chance.

Conclusions:

Based on the experiences described above, we have drawn up a set of general recommendations for implementing an RCT study:

- (1) In order maximize the chances of obtaining buy-in for the study, do in-person recruiting visits at all levels of school administration (district administrators, school principals and teachers), even if only district-level approval is needed to secure school participation in the study.
- (2) Factor in potential implementation challenges (due to weak implementation, threats to the integrity of random assignment or both) into the study design and statistical power analyses by oversampling the number of classrooms and schools needed to achieve the desired minimum detectable effect size.
- (3) Following recruitment, involve district administrators, principals and teachers in substantive discussions about the implementation of study and the protocols for service delivery as early as possible incorporate their feedback into the design of the study.
- (4) Continue communicating with sites and participants frequently throughout the course of the study, long after agreement to participate has been obtained.
- (5) Throughout the course of the study, establish a separate identity from the program you are evaluating.
- (6) Accompany experimental studies with equally rigorous qualitative studies that illustrate and explain why a policy intervention or instructional practice did or did not yield benefits downstream.

Our final recommendation is more general and less about the implementation of random assignment evaluations than about the types of interventions that are appropriate to be studied using a random assignment design. We would argue many of the interventions being evaluated with random assignment designs are ready for efficacy studies, but are not yet ready for effectiveness studies. The random assignment design requires that the program have a research base indicating association with the desired outcomes. It also requires that the program developers have the capacity to implement the intervention on a large scale. We postulate that there many of the interventions being evaluated in the RCT studies did not meet these criteria prior to evaluation.

Appendices

Appendix A. References

Burtless, G. 2002. Randomized field trials for policy evaluation: Why not in education? In F. Mosteller & R. Boruch (Eds.) *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Cook, Thomas D. 1999. *Considering the Major Arguments Against Random Assignment: A Analysis of the Intellectual Culture Surrounding Evaluation in American Schools of Education*. Paper presented at the Harvard Faculty Seminar on Experiments in Education. Cambridge, MA: March 1999.

_____. 2002. Randomized Experiments in Education: Why Are They So Rare? *Educational Evaluation and Policy Analysis*, 24 (3), 175-99.

_____. 2003. Why have Educational Evaluators Chosen Not to Do Randomized Experiments? *The Annals of the American Academy of Political and Social Science*, 589 (1), 114-149.

Cook, T & Payne, M. 2002. Objecting to the objections to using randomized trials in education research. In F. Mosteller & R. Boruch (Eds.) *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Condelli, L & Wrigley, H. 2004. *Instruction, language, and literacy: What works study for adult ESL literacy students*. Washington, DC: Office of Vocational and Adult Education.

Erickson, F., & Gutierrez. 2002. Culture, rigor, and science in educational research. *Educational Researcher*, 31 (8), 21 –24.

Gallagher, J. 2002. What's next for OERI? *Education Week*, 21 (28) 52.

Gueron, J. 2000. *The Politics of Random Assignment: Implementing Studies and Impacting Policy*. New York, NY: MDRC Working Paper #45.

Gueron, J. 2002. The Politics of Random Assignment: Implementing studies and affecting policy. In F. Mosteller & R. Boruch (Eds.) *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

McHugh, M., Gelatt, J., & Fix, M. 2007. *Adult English language instruction in the United States: Determining need and investing wisely*. Washington DC: Migration Policy Institute.

Mosteller, F., & Boruch, R. 2002. *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

National Research Council. 2002. *Scientific research in education*. Committee on scientific principles for education research. Shavelson, R. J & Towne, L., Editors. Center for education;

Division of social and behavioral sciences in education. Washington, DC: National Academy Press.

Olson, D. 2004. The triumph of hope over experience in the search for “what works”: A response to Slavin. *Educational Researcher*, 33 (1) 24 – 26.

Slavin, R. 2002. Evidence-based educational policies: Transforming educational practice and research. *Educational Researcher*, 31 (7), 15 –21.

St. Pierre, E. 2002. “Science” rejects post-modernism. *Educational Researcher*, 31 (8), 21 –24.

Stafford, B., Greenberg, D., and Davis, A. 2002. *A Literature Review of the Use of Random Assignment Methodology in Evaluations of U.S. Social Policy Programmes*. London: Center for Research in Social Policy.

US Congress. 2001. *No Child Left Behind*. Washington DC: Author.

Vinovskis, M. 2002. Missing in practice: Development and evaluation at the US Department of Education. In F. Mosteller & R. Boruch (Eds.) *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Wolf, P. 2003. *Looking Inside the Black Box: What School Factors Explain Voucher Gains in Washington, D.C.?* Paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA

Weiss, C. 2002. What to do until the random assigner comes. In F. Mosteller & R. Boruch (Eds.) *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Yap, Kim. O. et. al. 2005. *Can Experimental Research Be Conducted with Culturally Based Education Interventions? An Assessment of Feasibility*. Paper presented at Annual Meeting of the American Educational Research Association (AERA), Montreal, 2005.

