

RANDOMIZED CONTROLLED TRIALS AND PHD LEVEL TRAINING IN EDUCATIONAL RESEARCH

Robert Boruch and Erling Boe
University of Pennsylvania, United States of America
robertb@gse.upenn.edu

This invited paper reviews recent initiatives in teaching about randomized field trials at the graduate level in education research. Reports in Mosteller and Boruch (2002) are used as a benchmark. The initiatives have been driven heavily by recent governmental emphasis on “evidence based policy” in education, criminology, welfare and other sectors. This policy has been backed, in the US at least, by substantial investments in mounting randomized controlled trials to evaluate curriculum packages and programs in grades K-12, crime prevention programs, welfare, and other work. Accelerated growth in randomized trials has entailed cross discipline pre-doctoral and post-doctoral institutes and graduate education programs aimed at enhancing the quality in trial design, execution, and analysis of results. Specific illustrations of graduate education in this area of statistics are taken mainly from experience at the University of Pennsylvania’s Graduate School of Education and selected other institutions.

INTRODUCTION

The Mosteller and Boruch (2002) book, used as a benchmark here, stemmed from a state of the art conference undertaken at Harvard University on the topic of using randomized controlled trials to estimate the effect of education programs. The conference was stimulated by other events, as one may expect, including the earlier conduct of randomized trials. For instance, the state of Tennessee had mounted a large trial to produce estimates of the effect of reducing class size on the achievement level of students in the state’s public schools, randomizing teachers to classrooms and students to large or small class sizes. This Tennessee Class Size Experiment was a milestone in education, akin to the import of the Salk vaccine trials and the Streptomycin trials in health, in demonstrating the feasibility of such trials and the usefulness of results. Multiple independent analyses of the resultant data, by Mosteller and others confirmed original analyses in showing positive effects of reduced class size on children’s academic achievement. Information about this trial and others that have been exceptional, and evidence about their emerging importance, is covered in Mosteller and Boruch (2002).

The Education Sciences Reform Act of 2002 established the Institute of Education Sciences (IES) (<http://www.ed.gov/offices/list/ies>). By law, the IES is required to invest in education research that meets high *scientific* standards. This was to get well beyond research of the 1970s – 1990s which often (but not always) failed to meet such standards despite publications in peer reviewed journals. It was intended to change the culture of education research. IES succeeded in doing so in more than a few respects under the leadership of its director Grover (Russ) Whitehurst. This included remarkable support for large scale cluster randomized tests of education interventions. Nearly 80 such trials, for instance, have been mounted since IES’s creation, with 50-300 schools being involved in each trial.

These activities generated some controversy. In particular, portions of the education research community decried the use of randomized trials, arguing that education is more subtle than medicine, which depends on such trials for evidence, and that initiating programs of trials would be unwarranted on legal, ethical and pedagogical grounds. The National Academy of Sciences report on the topic (Shavelson & Towne, 2001) attempted to allay concerns of some parts of the education research community that a large scale investment in trials would swamp other research efforts, ethnographic case studies or surveys for instance, and to introduce balance into debates.

GRADUATE STUDIES IN EDUCATION RESEARCH

Prior to 2000, most programs of graduate study in education put little emphasis on randomized controlled trials in the field except perhaps as an aspiration. Indeed, much of the training based on much of the research was arguably poor in this respect. See Vinovskis’ paper in

the Mosteller Boruch volume. At the better institutions, including Penn, Northwestern, and others, one could take courses in statistical design of experiments and analysis of variance, for instance. Actual field experiments were few, however, and many such courses relied on examples from laboratory work or small scale studies, in educational psychology, for instance. Many good institutions required courses oriented toward observational data, i.e., surveys or administrative records, which continue to be far more common.

Once the federal emphasis on randomized controlled trials became evident and substantial funding was made available through IES and private foundation grants, changes occurred in a variety of ways. The changes were made in less than a decade at leading academic institutions. This also led to changes in orientation of for profit and on profit firms that do such research (often in collaboration with academic institutions).

GRADUATE STUDIES IN EDUCATION AT PENN: CORE PROGRAM COURSES

Penn's Graduate School of Education serves 100 PhD students each year. Partly as a consequence of increasing emphasis on evidence in general and quantitative evidence in particular, about 25% percent of these students are required to have a course in statistics regardless of their specialization. This requirement, lead by the Dean and approved by faculty, represents a shift and has caused some temporary dismay among students, in Teacher Education or Education Leadership for instance, who thought they might escape statistics in these specializations as they might escape statistics at other education schools. As one Harvard educated full professor declared snidely, in a discussion about students' difficulty or with or antipathy toward statistics, "We've all been there." This kind of professorial ignorance or arrogance is being superceded by societal interests in better evidence based partly on randomized trials.

At Penn, the Policy Research, Evaluation, and Measurements program (PREM), directed by the authors of this paper, is a major locus for graduate training on randomized controlled trials. The stress is on policy research, not policy per se, i.e. on evidence generated by research that informs policy. "Evaluation" and "Policy Research" in the title of the program includes statistically oriented courses that bear on estimating effects of programs in randomized trials, for instance, not on analysis of political policy.

To the extent possible, PREM core courses include those which we might expect: measurement, applied linear models, analysis of variance, survey design, and statistical uses of administrative records, and others. The trials-related courses depend on excerpts from classic texts such as Box, Hunter and Hunter (1978) and Kempthorne (1952/1983), as well as on newer work by Raudenbush and Bryk (2002) on hierarchical designs and cluster randomized trials in education and Boruch (1997) on non-statistical aspects in design, execution, and analysis of controlled trials. Measurement in this context includes Item Response Theory and related topics. Applied statistics journals such as *JASA*, *Journal of Educational and Behavioral Statistics* and *Journal of Research on Educational Effectiveness* carry pertinent information and are routinely used.

The PREM PhD program requires an orderly sequence of about 15 statistics courses applied in education contexts and further requires students to take 2-4 substantive matter courses so as to assure that the students are well informed about the education environments in which trials are mounted. For instance, one can trace a research and development cycle that engenders the use of statistical methods ranging from probability surveys of the nature and severity of a problem, such as school drop outs or teacher attrition, through development of interventions that are purported to resolve the problem and the measurement systems that they then require, such as conditional income transfer programs or ratings systems for teachers, and to the deployment of large scale field trials in which large samples of schools or classrooms or teachers are randomly assigned to different interventions to understand what works better to reduce school drop-out rate or to keep good teachers in schools. Examples from the US, Mexico, India, and elsewhere ar in ample supply and are used in the PREM course of studies.

Research internships have been a required part of graduate training at both MS and PhD levels in PREM since the 1990s. Theses internships usually involve collaborative work with faculty. As of 2000, PREM has put more emphasis on work off site on research projects, including participating in randomized trials, secondary analyses of data from randomized trials, and methodological studies (e.g. on missing data) in the context of trials. Internships opportunities are

driven by the grants that the statistically oriented faculty members have from IES and other funders.

All of the eight PREM faculty members, and all of the associated Adjunct members, have been directly involved with randomized controlled trials or statistically-oriented policy research in education over the last decade. See Penn's web site (<http://www.upenn.edu>) for credentials of Boe, Boruch, Fantuzzo, MacDermott, and Porter, and of Turner, May, and Victor. All of the PREM Penn faculty have received recognition for their work in one form or another. This list includes an elected Fellow of the American Statistical Association and the American Academy of Arts and Sciences, a former president of the American Educational Research Association and a member of the National Academy of Education, and recipients of awards from diverse professional organizations including Head Start and NCEA. Each faculty member, however, is constrained to admit only one PhD applicant each year who is then fully funded through a fellowship.

Many randomized controlled trials in education involve interdisciplinary teams. This provokes cross fertilization of ideas. Graduate students in PREM have, since the 1990s, been encouraged to engage in courses across departments and schools at Penn. This emphasis became more emphatic and better funded as IES began to fund interdisciplinary fellowships in education research. Most PREM students capitalize on the opportunity, as part of their ordinary options in PREM's course of studies. They may learn from Rosenbaum, in Wharton's Statistics Department, about observational studies, for instance. They may learn from Allison, in the Sociology Department, about survival analysis and missing data in the context of trials. They may learn about trials in other contexts from Berk in Criminology, Strom in Biostatistics, among others. The aim is to assure that students understand that the goals and issues engendered by randomized controlled trials transcend discipline boundaries but that some issues are tied to particular contexts. For instance, using schools as the unit of random assignment versus using police "hot spots" as the units in their respective contexts requires that one recognize different constraints on the relevant social and institutional systems, different incentive structures, and so on so as to tailor the experiment's design to suit the setting.

Thirteen high end academic institutions that have been funded by IES over the last eight years have a similar interdisciplinary approach; they are identified here in different aspects of the effort. See <http://ies.gov/ncer/projects/program.asp?ProgID=16>. Obviously, PREM-like programs at Penn and elsewhere, are in the minority. That is, most schools of education including graduate schools do not put much emphasis on statistical methods, much less on randomized trials. The press in such institutions toward qualitative information, policy and practice, and theory (as opposed to empirical field tests) is still very high, and often swamps statistical interests. This tilt was, to be fair to education research people, also evident in the early history of economics and in medicine.

BEYOND A CONVENTIONAL CORE COURSE OF STUDIES

A well organized program of studies based at a university, however excellent the university may be, is insufficient for graduate studies and for post doctoral education. The shift toward large scale complex field tests of education and other social programs has forced PREM faculty at Penn's Graduate School of Education, and other colleagues at Florida State, Vanderbilt, Michigan, for instance, to think beyond conventional graduate coursework, internships, and graduate assistantships.

Funded Lecture Series and Pre-doctoral Programs

Between 2004 and 2008, the Institute of Education Sciences (IES) funded 13 pre-doctoral and 17 post-doctoral fellowships for cross-discipline work in education. In PREM at Penn, and selected other places, the pre-doc program has been well integrated with a guest lecture series in which guests travel is paid and guests receive a speaker's honorarium. The funding for such activity at Penn was absent during the 1980s and 1990s, though PREM's Directors Boe and Boruch managed often to engage interesting people who further educated graduate students.

To routinely engage external experts routinely and at a reasonable scale, funding from agencies such as IES must be sought. Penn faculty colleagues, such as Rebecca Maynard and Laura Desimone, with other faculty members' support, have been assertive and successful in

obtaining such funding. At Penn's Graduate School of Education, the lecture series is attended by IES PhD student Fellows, and by faculty and graduate students from different academic departments. The guest lecturers have included senior people in applied statistics such as Larry Hedges (Northwestern), as well as junior and mid level people doing quantitative education research in economics, child development, evaluation programs, and child development. Guests from research firms who are engaged in large scale field experiments such as Turner (Analytica), Schochet (Mathematica), Garet (American Institutes for Research), and Aber (New York University) has been invaluable in keeping abreast of efforts to better design trials so as to meet field conditions and scientific needs.

Off Site Conferences and Workshops: Foundations and Agency Sponsored Efforts

At least some efforts early in 2000-2002 helped to jump start senior researcher's expertise in large scale trials. The Harvard University conference that led to the Mosteller Boruch (2002) publication is one example. Beyond this, the Rockefeller Foundation sponsored a cross discipline set of meetings on the topic of place randomized trials in Bellagio Italy, to develop state of the art papers on such trials in education, health, crime, and welfare. Products of the work appeared in a special edition of the *Annals of the American Academy of Political and Social Sciences* (Boruch, 2005). The Bellagio meetings served as a platform for subsequent conference in New York that engaged about 50 senior people including allied statisticians such as Steve Raudenbush and Alan Donner with senior quantitative researchers from different disciplines.

The William T. Grant Foundation (<http://www.wtgrantfoundation.org>), under the eldership of its president Bob Granger, built on these efforts to develop workshops for senior education researchers with a strong interest in learning about advances in statistical methodology and trials. The first workshop, undertaken at the University of Michigan, for instance, focused on advances in theory, methods, and software, including power analysis, for hierarchical (multi-level) setups. It covered setups in which entire schools or classrooms are the units of random allocation to different interventions and teachers or students, nested within these units, may themselves be randomly assigned to interventions.

The Institute of Education Sciences (IES) followed suit with a series of workshops that broadened the possible participation. With IES support, Northwestern University and Vanderbilt University, for instance, laid on workshops that included graduate students, faculty from other institutions, and young researchers from non profit and for profit organizations with an interest in learning about such trials. PREM faculty contributed to some of the workshops. And current and former PREM PhD students often attend them at the encouragement, and at times with financial support, of PREM faculty members.

New Societies, Collaborations, Organizations, and Journals

New scientific societies are formed as science becomes more complex and as specializations within the science develop. This occurs, of course, in the physical sciences. The American Association for the Advancement of Science, for example, is an old and revered organization that covers all of the sciences in principle. On account of the need for specialization, however, other organizations, such as the American Institute of Physics, have developed to service the needs of individuals with more specialized interests and skills. One can see similar phenomena with the statistical sciences, e.g. the invention of the Psychometric Society, the Society for Multivariate Research, and others.

At least two new societies have been formed since 2002 to direct attention to higher quality statistical work in estimating effects of interventions, especially randomized trials, in the education sciences and other social sciences. The Society for Research on Educational Effectiveness (SREE), for instance, runs annual meetings, often coupled with workshops, and publishes the *Journal of Research on Educational Effectiveness* so that one may continue to learn. The Society's Board includes representatives that have benefitted from IES funding for trails oriented pre-doctoral work including Penn, Northwestern, Berkeley, and Florida, as well as from like minded research firms such as American Institutes for Research and MDRC. The Academy for Experimental Criminology was created to foster randomized trials in the area, e.g. on police

interventions in crime hot spots. The *Journal of Experimental Criminology* is a major new source of information on trials in this arena. Faculty and PhD students capitalize on these, of course.

The international Campbell Collaboration, formed in 1999, directs attention to systematic reviews of research on effects of interventions in crime and justice, education, and social welfare (<http://www.campbellcollaboration.org>). High value is attached to assembling and reviewing findings from randomized controlled trials. The Cochrane Collaboration (<http://www.cochrane.org>), older sibling of Campbell, was formed in the early 1990s and also attaches high value to randomized trials. Moreover, both Collaborations exploit and help advance the state of the art in combining estimates of effect, including meta-analysis. The statistical pedigree in the meta-analytic work dates back to statisticians William Cochrane, Frederick Mosteller, and Ingram Olkin. Indeed, the Campbell Collaboration gave Olkin the Mosteller Award for Olkin's contributions to the area (the award, of course, is named after Mosteller).

Information about these resources and their activity in regard to randomized controlled trials are routinely incorporated into courses for PREM students at Penn and elsewhere. Virtually all related faculty members, in PREM for instance, routinely forward email messages and notices to the list serve roster of graduate students so that the students are aware of opportunities to learn and to actively contribute to such meetings in jointly authored papers, poster sessions, and in other ways. Graduate students have routinely attended meetings of these organizations and, when possible, their travel expenses are covered.

Actualizing Randomized Trials in Context: Major Non-Statistical Constraints and Options

The idea that a well run randomized trial results in a fair (statistically unbiased) estimate of an education program's effect and a legitimate statistical statement of one's confidence in the results is fundamental. At Penn's Graduate School of Education, the PREM program tries to assure that graduate students understand that this idea must be supported by high skills in non-statistical arenas.

For example, PREM's most advanced seminar on randomized controlled trials includes consideration of the (a) managerial, (b) political-institutional, (c) ethical, and (d) scientific issues that affect the design of a controlled trial, and the trial's execution. The origin of this particular course lies in earlier works on social experiments sponsored by the National Science Foundation, notably Riecken et al. (1974) and Boruch (1997), and in the papers in Mosteller and Boruch's (2002) book.

To put the matter bluntly, the students then come to understand this: Unless the prospective participants in a randomized trial trust the statistician, the trialist, the trial will not be done. Understanding how to build partnerships in the social sector that permit a trial is no easy matter. The advanced PREM seminar covers contemporary solutions to political-institutional and ethical issues that randomized trials may engender, and attends to the science underlying the rationale for a trial. This particular seminar exploits resources beyond the instructor of record, Boruch, in ways described below.

To the extent possible, guest lecturers who have succeeded in resolving issues or reducing obstacles in a trial are engaged. For instance, guest lecturers have included people from schools who have influence over whether a trial will be run in the sense that they can enhance the cooperation of teachers, parents, and students in a trial on the effects of a curriculum package. Because many trials are mounted by organizations apart from universities, the relevant web sites are used. For instance, material from continuously updated web sites of the Manpower Research Development Corporation (MDRC), American Institutes for Research (AIR), Mathematica Policy Research, the International Initiative for Impact Evaluation (3ie), and others is used routinely.

Access to Information: Internet

Apart from the web sites for organizations that design and execute randomized controlled trials, and aside from similar web sites for new organizations identified earlier that consolidate professional interests in randomized trials in education and the social sciences, access to pertinent information has advanced remarkably over the last eight years.

The William T. Grant Foundation, for instance, has sponsored the posting of micro-records from the Tennessee class size experiment for instance. This access has permitted a variety of senior

people, including Mosteller, to reanalyze the data. Young colleagues, such as Spyros Konstantopolis (reference), capitalized on these data from the Tennessee class size experiments to produce original and interesting work while a graduate student at Northwestern University and as assistant professor at Boston University. Web access to the micro-records from randomized trials in education research is not yet common. Nonetheless, the precedent set by the Grant Foundation is important. It can be used as a model of what should happen after the trial is completed and the original analysis is done, so as to permit independent confirmation of results, testing new hypotheses or theories, and methodological studies on ways to analyze good but imperfect data.

The Grant Foundation also sponsors and posts advanced software for sophisticated power analyses in the context of cluster/group/place randomized controlled trials. The web accessible software, called Optimal Design, along with ample documentation and references to the mathematics that underlie the software operations, are used routinely in relevant PREM courses on trial design and relevant PhD courses at other institutions. At the University of Michigan, for instance, Jessaca Spybrook's dissertation entailed acquiring the research design papers for nearly 60 such trials from as many trial teams, and then reanalyzing the work with special attention to exploiting new developments in statistical power analysis.

Private foundations can catalyze intellectual initiatives, as the W. T. Grant Foundation (<http://www.wtgrantfoundation.org>) has in the arena of randomized trials. Governments can do far more than a private foundation and do so, at times. The What Works Clearinghouse (WWC), for instance, was created by the US Institute of Education Sciences (<http://whatworks.ed.gov>) as a source of dependable scientific information about effective programs in education. Randomized controlled trials are put at high priority in judging effectiveness of interventions reviewed by the Clearinghouse. Inasmuch as these trials are imperfect, the WWC web site regularly carries state of the art documents on statistical advances that inform its reviews. The PhD students at Penn and elsewhere can then learn about how decisions might be reached on trial results when attrition from original randomized samples is high and when attrition rates differ in the arms of the trial. They learn from these state of the art documents about orders of magnitude of intra-class correlations that influence estimates of variance and formal and conventional statistical tests of a null hypothesis. They learn also about how to put the statistical notion of an effect size, a difference between means divided by the square root of a pooled variance within the arms of the trial, into common parlance for teachers or parents.

REFERENCES

- Boruch, R. F. (Ed.) (2005). Place Randomized Trials: Experimental Tests of Public Policy. *Special Issue. Annals of the American Academy of Political and Social Science*, 599, (Entire Issue).
- Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for Experimenters*. New York: Wiley.
- Kemphorne, O. (1952/1983). *The Design and Analysis of Experiments*. New York: Wiley/Malabar Florida: Robert Krieger Publishing.
- Mosteller, F., & Boruch, B. (Eds.). (2002). *Evidence Matters: Randomized Trials in Education Research*. Washington, DC: Brookings Institution Press.
- Neter, J., Kutner, M. H., Nachsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models*. New York: McGraw Hill.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks Ca: Sage Publications.
- Riecken, H. W. et al. (1974). *Social Experimentation*. New York: Academic Press.
- Shavelson, R., & Towne, L. (Eds.) (2001). *Scientific Inquiry in Education: Report of National Research Council Committee on Scientific Principles in Education Research*. Washington, DC: National Academy of Sciences Press.