

## *Evidence Matters. Randomized Trials in Education Research* by Frederick Mosteller and Robert Boruch

The edited volume *Evidence Matters* (Brookings Institution Press: Washington, DC, 2002) is an extensive debate on the merits of using randomized trials in education research. The stakes in the debate are high, because a shift toward experiments would significantly alter education research across many dimensions, such as professional training, funding opportunities, labor intensity, and publication strategies. This volume mainly focuses on the methodological reasons for, and implications of, randomized trials, but the contributors are sufficiently diverse and offer distinct perspectives on the practicalities of research.

Judith Gueron ('The Politics of Random Assignment') details her considerable experience in implementing social experiments. This Chapter includes numerous pithy maxims about research out in the field, and the result is a useful handbook or – more appropriately – checklist of problems (not only political but also practical). It would have been interesting to know whether and to what extent the problems were solved, rather than impairing the quality of the research. But the Chapter ends with numerous cautions, particularly for research in education.

Robert Boruch and co-authors argue in favor of randomized field trials (RFTs) in education. They set out five standards by which to assess the importance of RFTs: do they answer important questions? Have they been important in the past? Are they ethical? Are they being undertaken? And can they be scientific? It is the last of these that is the most contentious. Although the methodological advantages of RFTs are compelling, there may be practical disadvantages: the practical differences between research from RFTs and non-RFTs are discussed here, but the conclusions are decidedly cautious. The differing results across research methods cannot simply be explained by the weaker protocols used in non-RFTs.

In their discussion of the relationship between resources and outcomes, David Cohen and co-authors change the direction of this volume somewhat. They offer a choice between research agendas where the key problem is either “to identify and then deploy the resource mix most likely to improve learning” or “to identify and mobilize the knowledge, practices, and incentives that will enable them to best use resources”. Whereas the first agenda might draw directly on experimental evidence, the second agenda might suggest systemic reforms to newly align incentive structures. Chapter 5 by Maris Vinovskis also sits rather incongruously in the middle of this book. It charts the development of evaluation at the U.S. Department of Education, with a title ‘Missing in Practice’.

The most interesting Chapter is that by Thomas Cook and Monique Payne on ‘Objecting to the objections to using random assignment in educational Research’. Nine objections are interrogated, and each is addressed directly and intelligently. Cook and Payne are also prepared to concede that some of the objections may be valid in some cases. So, to one objection – experiments suggest a oversimplified theory of causation – Cook and Payne are sympathetic to the contention that “biased answers to big explanatory questions are more important than unbiased answers to smaller causal-descriptive questions” (p.154). Many of the objections, however, reduce to statements either that the social world is more complicated than the physical world or that random assignment is not practical (e.g. ethically; similar arguments are made in the subsequent, brief Chapter by Gary Burtless, who considers the political antipathy to RFTs in education).

The first statement cannot easily be dismissed, and the second is obviously context-dependent. Even if there is case for more experimental research in education in general, the case also needs to be made in the particular. Explicitly, Cook and Payne reject the triumphalist notion of experiments as a ‘gold standard’: “It is clearly not that in educational contexts, given the

difficulties with implementing and maintaining randomly created groups, with the sometimes incomplete implementation of treatment particulars, with the borrowing of some treatment particulars by control group units, and with the limitations to external validity that often follow from how random assignment is achieved” (p.174). In favoring random assignment, Cook and Payne emphasize experiments' high internal validity and their yielding of a counterfactual.

The final Chapter, by Carol Weiss, asks ‘What to do until the random assigner comes?’ Her response, drawing on Theory-Based Evaluation, is essentially ‘Specify your hypothesis as tightly as possible’. Although this response is a good one, it only addresses concerns over internal validity and so might be complemented by ‘Specify to what use your results might be put (again as tightly as possible).’

Overall, *Evidence Matters* is a very useful deliberation on the experimental method. Its main conclusion is undeniable: RFTs have high causal integrity for the intervention being tested. But, the contributors make clear two caveats should be attached. First, RFTs need to be implemented properly and this is no simple matter. Second, the decision to employ random assignment should be a considered one: research designs should be chosen based on the costs and benefits of each method, and these vary with the circumstances. To these might be added a third caveat, regarding weak external validity. Indeed, the problems of inference from observation abound: a reader expecting a book called *Evidence Matters* to make the case that evidence matters would be in error; this *Evidence Matters* is about how we should collect the evidence that matters. There is a difference.

**Clive R. Belfield**

[cb2001@columbia.edu](mailto:cb2001@columbia.edu)

National Center for the Study of Privatization in Education, Teachers College, [www.ncspe.org](http://www.ncspe.org).