



Expert Opinion

An evidence-based policy?

By JONATHAN SUOVITZ

The No Child Left Behind Act has sparked a fiery debate within the educational research community about the role of experimental evidence in educational research. Yet, this debate sidesteps two larger questions. First, how can we improve the quality of educational evidence overall? And second, are the consumers of educational evidence prepared to distinguish between different qualities of evidence? By restricting the range of what is considered legitimate evidence, and without concomitantly sharpening the discernment of local policymakers to make use of better evidence, we may find ourselves with more definitive answers

to a narrow range of peripheral questions and a field ill prepared to make use of them.

QUALITY EVIDENCE

Federal policymakers are appropriately devoted to the idea that evidence produced by educational research studies that examine the effectiveness of a specific program or educational intervention should be “scientifically based.” Scientifically based research as defined in NCLB are studies that apply “rigorous, systematic, and objective procedures” subjected to peer review. What is controversial is the clear preference for randomized experiments and the ensuing perception that these are the only legitimate source of evidence about programmatic effectiveness.

Randomization produces the most definitive and precise estimates of effect because it rules out alternative explanations. Researchers at MDRC—a social policy research organization based in California and New York—for example, have conducted comparative studies of randomization versus quasi-experimental designs. They found that even the most sophisticated analysis techniques in quasi-experiments do not come close to the estimates from randomization. As a statistics professor once told me, “You can’t fix with analysis what you have bungled in design.”

That doesn’t mean we need only experimental studies all the time. Case studies, for example, may be the place where causal hypotheses are first generated. There are other interventions that are not ready for randomization. For example, many technology initiatives are not yet well developed enough to merit the expense of randomization. Furthermore, there are other big and important questions that must rely on “natural experiments.” For example, if we seek to understand the effects of state graduation policies on high school dropout rates, it is not feasible to randomly assign graduation policies to states. In other situations, there may be ethical concerns that preclude randomization. In short, because randomized studies are expensive, difficult to conduct, sometimes not feasible, and even unethical in some situations, they should only be used in particular circumstances.

Focusing the debate on one corner of a much larger terrain masks the larger problem of the quality of educational evidence. Too many educational studies, regardless of their specifics, are fraught with poor designs, meager measures, inappropriate analyses, and unsubstantiated conclusions. There is no clear articulation of what is quality research in the education field.

Several organizations have put

forth ways of thinking about this problem. But what is missing is widely distributed and accepted guidance for the larger field that can guide young researchers, journal reviewers, and decisionmakers alike. While there will always be gray areas and ongoing debates, there are general features—like representative samples, comparison groups, and effect sizes—that make some research clearly better than others.

The danger of undue focus on randomized experiments is that there is a potential for conducting better substantiated studies of less meaningful interventions. For example, it is far easier and cheaper to randomly assign students to after-school programs or short education seminars than it is to assign schools to comprehensive school reform programs. Yet several CSR programs are promising and powerful interventions that merit support for randomized trials. On the flip side, there are many interventions not ready for randomized experimentation, because they do not have quasi-experimental or even case evidence that they are promising.

CRITICAL CONSUMERS OF EVIDENCE

By calling for the adoption of only those practices that are scientifically

See EXPERT, Page 39

Expert

Continued from Page 40

based, the law is also creating a demand for using evidence in support of adoption decisions. Currently, this demand is artificial, propped up by governmental support for only those programs that can pass the “evidence-based” litmus test. The demand is not coming from the public school marketplace for programs, which currently has difficulty distinguishing between evidence and conjecture, much less differentially effective programs.

My colleague at the Consortium for Policy Research in Education, Henry May, and I recently completed a round of interviews with state, district, and school leaders about their understanding of evidential quality, the role of evidence in local decisionmaking, and their willingness to participate in randomized experiments. With some notable exceptions, most local policymakers had vague and simplistic notions of what makes strong evidence. Important qualities such as comparison groups, statistical tests of difference, and unbiased sampling were largely absent in their explanations of what makes strong evidence. We have the distinct impression that any savvy pitch featuring large samples, student demographics similar to theirs, and evidence of improved student

achievement (whether valid or not) could sell a program.

We also learned that, unlike its status in the rhetoric of the law and academic talk, evidence does not rank high on the list of criteria for program adoption decisions. Local decisionmakers prioritized compatibility and feasibility, including philosophical alignment and support for implementation before evidence of effectiveness. For example, one superintendent of a mid-sized urban district explained, “I choose a program first and foremost based upon its fit with our schools, and its philosophical alignment with what we are doing and only then . . . its evidence of improved student learning.”

Adoption decisions are also frequently political, which cuts against educators’ willingness to participate in educational research. This explains, at least in part, the overwhelming lack of enthusiasm for randomized experiments among our respondents. No matter how we framed proposed research using experimental conditions (lottery assignment, deferred treatment, cash compensation for those put in the control group), there were almost no takers. “Well, we’re not going to do that, so you can forget about that,” said one state education official. “Any superintendent that didn’t serve the neediest schools first would be out of a job in the morning.” An

urban district superintendent was equally disdainful. “There is no way in the world we would participate in a randomized experiment . . . We are not going to be the guinea pigs.” School principals would only participate if it was district imposed and they had no choice.

NCLB’s pressure for improvement also reduced willingness to participate in experimental studies. For all schools, and particularly the low-performing ones who are most often the target of research studies, the pressures to meet adequate yearly progress precluded taking a risk of being in a control group. More subtly, the request to conduct a study perpetuates the perception that there is no evidence about a program’s effectiveness. District and school leaders were only willing to consider participation in small studies in peripheral areas. One superintendent, for example, thought he could identify a handful of schools that might “as part of a summer program or extended day program” consider taking part. A district superintendent in another state said he wouldn’t advocate participation unless “. . . it was something that wasn’t a very significant program.”

In the 1930s, when government reformers sought to improve the quality and utility of medical research, they tackled the problem in two ways. First, through what

became known as the FDA, they clearly defined quality research in the field. They also realized that the value of the evidence depended upon the knowledgeable and skillful use of it, and they worked with professional associations and universities to train professionals to assess the evidence. The challenge the education field faces today is similar. We must improve the quality and rigor of all kinds of educational research. We must also educate the consumers of educational research so that they can separate the wheat from the chaff. Only when educational leaders have a sophisticated understanding of what scientifically based research is can they make truly informed decisions. ■

Jonathan Supovitz is an associate professor in the Graduate School of Education at the University of Pennsylvania and a senior researcher at the Consortium for Policy Research in Education.

POSTSCRIPT: For other viewpoints on the use of randomized trials in education research, read Robert Boruch’s “The Virtues of Randomness” and Thomas D. Cook’s “Sciencephobia” in back issues of Education Next, www.educationnext.org.