

Published Online: February 5, 2008 *Education Week*

Published in Print: February 6, 2008

Includes correction(s): February 7, 2008

Commentary

# **‘Scientific Research’ and Policymaking**

## **A Tool, Not a Crutch**

By Frederick M. Hess & Jeffrey R. Henig

These are heady times for education researchers. The No Child Left Behind Act famously endorses the use of “scientifically based research,” the federal Institute of Education Sciences has elevated the profile of rigorous scholarship, and presidential candidates tout studies on teacher quality, testing, and school choice. Advocates market favorable social science evidence and enlist sympathetic researchers as spokespersons. This attention can tempt researchers to oversell their findings and policymakers to overinterpret them—confusing our understanding of what “scientific research” can and cannot teach us when it comes to education policy.

We write as two individuals housed in very different institutions and frequently on opposing sides in polarized policy debates, both having just published books plumbing the impact of research on education policy. One sits in a school of education; the other in a Washington think tank typically described as “conservative.” Despite our differences, we share the concern that undisciplined claims about the power of research can stand in for careful thinking, foster cynicism, and undermine the long-term contribution of the research community.

There is a natural desire for education scholars to resolve complex questions by providing definitive answers, as their brethren have done in biology or chemistry. In practice, the hope that a “killer study” will settle vexing policy questions regarding merit pay or charter schooling gives rise to heated, anxious, and very public cycles of attack and counterattack. The result is good theater and a potent fundraising opportunity for advocates, but also a diminished appetite for findings that muddy partisan debates, and less opportunity for autonomous researchers to serve as a credible check on overblown or dubious claims.

We offer the seemingly paradoxical admonition that effectiveness and professional responsibility counsel that researchers should promise less when it comes to determining what policies “work.” In doing so, we are adamantly not calling for “niceness” or consensus—we believe it appropriate that meaningful policy disputes feature sharp-edged disagreement among scholars keeping a skeptical eye on one another’s methods and interpretations.

We are, instead, advocating common sense and humility about what research can provide. Generally absent from debates over methodologically complex and technical work “proving” that licensure does or does not improve teacher quality, or that charter schools do or do not raise test scores, is an appreciation of what the “scientific method” can and cannot deliver. In many of these debates, research is unlikely to provide the definitive judgments that policymakers seek. Three insuperable obstacles are responsible.

First, education metrics—no matter how seemingly precise—are almost inevitably at least one step removed from the concepts that we are really interested in. Reading and math scores are only an

approximation of learning. Eligibility for free lunch is a crude proxy for family income. Some data that would be enormously helpful—student-level information about achievement and family background collected in comparable form across districts and states—is legally and politically difficult for governments to collect because of privacy concerns and the design of federalism.

Generally absent from debates over methodologically complex and technical work is an appreciation of what the 'scientific method' can and cannot deliver.

A second problem entails generalizing across place and time. Details about policies and the contexts in which they are implemented vary across locales. Just as an employee pay system that has been shown to work for Google will not necessarily work for Citigroup or Stanford, so too a study that convincingly shows a program works in one setting cannot casually be presumed to yield the same outcomes elsewhere. Meanwhile, policies may take time to mature, or early success may be due to the skill of early adopters, enthusiasm associated with hot ideas, and foundation support . . . but then peter out. Researchers warning that it is “too soon to tell” about a policy may try the patience of policymakers who need to make decisions today, but that does not make their cautions any less justified.

Third, the most vexing challenge is that of making causal inferences in a complex social world. Randomized field trials are the research design of choice precisely because of their potential to precisely establish cause and effect. That is why randomized clinical trials serve as the “gold standard” in medical research. Efforts to adopt the “medical model” in schooling, however, have been plagued by a flawed understanding of how the model works in medicine and translates to education.

The medical model, with its reliance on trials in which drugs or therapies are administered to individual patients under explicit protocols, is enormously powerful and prescriptive when recommending interventions for discrete medical conditions, but few imagine it authoritative when considering the merits of universal health coverage or how best to hold hospitals accountable. While the federal Food and Drug Administration monitors and approves drug therapies, its approval is not required before a hospital alters hiring practices or accountability metrics. In fact, when it comes to questions of governance, accountability, and compensation, few would suggest the health sector is faring a whole lot better than is education.

In education, randomized field trials are the optimal course for assessing pedagogical and curricular approaches for increasing knowledge and skills via the application of discrete treatments to identifiable students under specified conditions. Such interventions are readily susceptible to randomized field trials, which yield results that can reasonably serve as the basis for prescriptive policymaking.

But leading policy controversies are not about those things. Organizational reforms relating to governance, management, compensation, and deregulation are rarely precise and do not take place in controlled circumstances. Research can illuminate the impact of such reforms and how context matters, but is unlikely to determine with any surety whether given policies “work.” Much as we may wish it were otherwise, research into topics like merit pay or decentralization will always be more useful as a proximate guide than as a basis for prescriptive policymaking.

Undisciplined claims about the power of research can stand in for careful thinking, foster cynicism, and undermine the long-term contribution of the research community.

Let us be perfectly clear: Good research has an enormous contribution to make—but, when it comes to policy, this contribution is more tentative than we might prefer. Scholarship’s greatest value is not the ability to end policy disputes, but to encourage more thoughtful, disciplined, and tempered debate. In particular, rigorous research can establish parameters as to how big an effect a policy or program might have, even if it fails to conclusively answer whether it “works.” For instance, quality research has quieted

the notion that either Teach For America recruits or national-board-certified teachers are likely to have heroic impacts on student achievement.

Researchers can also challenge orthodoxies and raise questions about the conventional wisdom—as has been the case with scholarship questioning the relationship of spending to school performance, or between teacher credentials and teacher performance. Finally, research can illustrate uncomfortable realities, as with research documenting the allocation of dollars in urban districts, the extent of the dropout problem, or the impact of collective bargaining on district practices.

While individual studies cannot carry the weight of contemporary expectations, research as a collective enterprise—the clash of competing scholars attacking questions from different perspectives, with varied assumptions and methodologies—can leave us wiser and better equipped to make good choices. Ultimately, though, sifting through the accumulated evidence can only inform decisions; it cannot relieve us of them.

Frederick M. Hess is the director of education policy studies at the American Enterprise Institute and the editor of *When Research Matters*, published this month by Harvard Education Press. Jeffrey R. Henig is a professor of political science and education at Teachers College, Columbia University, and the author of *Spin Cycle*, co-published by the Russell Sage Foundation and the Century Foundation in February.