



## When Education Research Matters

By Frederick M. Hess

*The tangled relationship between education research and policy has received little serious scrutiny, even as paeans to “scientifically based research” and “evidence-based practice” have become a staple of education policymaking in recent years. For all the attention devoted to the five-year-old Institute of Education Sciences, to No Child Left Behind’s (NCLB) call for “scientifically based research,” to professional interest in data-driven decision-making, and to the refinement of sophisticated analytic tools, little effort has gone into understanding how, when, or why research affects education policy. Instead, most discussion has focused on how to identify “best practices” or “scientifically based” methods and how to encourage classroom educators to use research findings. In *When Research Matters: How Scholarship Influences Education Policy*, a new book published by Harvard Education Press, Frederick M. Hess has collaborated with a team of leading scholars to examine these questions.*

For nearly two decades, researchers and advocates have touted the idea of reducing class size to improve student performance as a no-brainer. University of Wisconsin professor Douglas Harris recently reported that 88 percent of parents support class size reduction and that teachers endorse it by a similar margin.<sup>1</sup> The American Educational Research Association advised in a 2003 policy brief that reducing class size should be a top funding priority among school initiatives.<sup>2</sup> But the reality is that most research findings are equivocal on the subject, and even those that are not should be handled with due caution when crafting policy solutions.

The push for class size reduction has benefited enormously from the findings of the famed Student Teacher Achievement Ratio project (STAR), a class size experiment conducted in Tennessee in the late 1980s. Educators spent \$12 million on STAR between 1985 and 1989 to examine the impact of class size on student learning.<sup>3</sup>

---

Frederick M. Hess (rhess@aei.org) is a resident scholar and the director of education policy studies at AEI. He is the editor of *When Research Matters: How Scholarship Influences Education Policy* (Harvard Education Press, 2008).

Researchers found significant achievement gains for students in small kindergarten classes and additional gains in first grade, especially for black students, which persisted as students moved through middle school.<sup>4</sup>

The STAR results suggested that this crowd-pleasing reform could ease teachers’ working conditions while also boosting student achievement. Not surprisingly, the research quickly found favor with and was trumpeted by teachers unions and advocates for increased school spending.

The STAR findings were applied famously and recklessly in California, where legislators adopted a class size reduction program in 1996 that cost \$771 million in its first year and \$1.7 billion annually by 2005.<sup>5</sup> The only major evaluation of California’s program, conducted by the American Institutes for Research and the RAND Corporation, found no impact on student achievement.<sup>6</sup>

What happened? First, policymakers were inattentive to the nuances of the findings and failed to take into account the considerations of a different setting. California’s initiative created an incentive for school districts to place first and

second graders (and, soon after, kindergartners and third graders) in classes of no more than twenty students. However, classes of twenty were substantially larger than those found effective by STAR, and the strategy was applied without STAR's narrow focus.<sup>7</sup>

Second, STAR was a pilot program—externally funded and directed to a limited population—and reformers did not account for the changed context in California. The benefits of class size reduction appear much harder to capture when the strategy is embraced by multiple schools drawing from a limited teacher pool. Widespread class size reduction created a voracious appetite for new teachers and diluted teacher quality.<sup>8</sup>

In practice, the findings from the much-heralded experiment turned out to be a far less useful guide for policymakers than many had hoped. In fact, apart from the STAR project, research on the merits of reducing class size shows mixed results. In 1999, Stanford economist Eric Hanushek reported that 277 econometric studies of student performance conducted through 1994 had examined the impact of class size or student-teacher ratios on achievement. Of those, just 15 percent found statistically significant positive effects, and 13 percent found statistically significant negative effects.<sup>9</sup> A slew of states, including Florida, Nevada, and Utah, have nonetheless pursued class size reduction in recent years, despite the uncertain evidence as to whether its benefits are commensurate with its costs.

## The Limits of “Scientific” Policy Research

The class size example highlights the degree to which research is not a purely technical endeavor but, rather, must be understood as part of an ecosystem of interpreters, advocates, funders, and policymakers. Efforts to emulate the medical community's effective use of randomized field trials are legion, but the enthusiasm for such a model is frequently accompanied by an imperfect understanding of its limitations and how it applies in education.

At the most basic level, especially in a context marked by fevered efforts to address the nation's achievement gap, this enthusiasm can lead to over-promising and unrealistic expectations of timelines and solutions. As acclaimed Harvard physicist Gerald Holton has observed, “Practitioners of science know well that the path is strewn with hurdles and pitfalls, [and] costly detours.”<sup>10</sup> More pithily, renowned biologist Stephen Gould has lamented, “Over 90 percent of the day's work generally turns out to be for naught, and then you still have to

clean out the mouse cage.”<sup>11</sup> The desire to identify interventions quickly that will take effect almost immediately, as with the urgent time horizons envisioned by NCLB-style accountability, generates a reluctance to accept the arduous realities of the scientific process and can favor glib researchers rather than diligent ones.

---

Efforts to adopt the “medical model”  
in education research have been  
plagued by a flawed understanding  
of how the model translates.

---

More fundamentally, randomized field trials are the research design of choice precisely because of their potential to 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 education research, however, have been plagued by a flawed understanding of how the model translates.

The medical model, with its reliance on trials in which drugs or therapies are administered to individual subjects under explicit protocols, is enormously powerful and prescriptive when recommending interventions for discrete medical conditions. Few imagine it as authoritative, however, when considering the merits of universal health care coverage or how best to hold hospitals accountable. While the Food and Drug Administration monitors and approves drug therapies, its approval is not required before a hospital alters management practices, compensation strategies, or accountability metrics.

In education, the medical model's reliance on randomized field trials is 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 and yield results that can reasonably serve as the basis for prescriptive policymaking.

Many of the biggest controversies in education, however, are not about pedagogy or curriculum but relate to governance, management, compensation, and deregulation. These policies are rarely precise and do not take place in controlled circumstances. Research can shed light on how such reforms unfold and how context matters, but it is unlikely to determine with any surety whether such 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 prescription for policymaking.

## The Democratization of Dissemination

Thirty years ago, it was unusual for academics to release their work directly to a policy audience; there was no convenient way to do so except to mail it to designated recipients. Today, of course, the Internet has fundamentally altered that calculus. The conventional approach to dissemination—in which scholars rely on academic conferences, professional associations, books, and scholarly journals to communicate their findings—has been challenged by the proliferation of national and state-based think tanks and advocacy groups and their successful use of inexpensive dissemination strategies.

---

Many of the biggest controversies in education are not about pedagogy or curriculum but relate to governance, management, compensation, and deregulation.

---

The transformation has opened discourse, raised new questions about how research quality can be ensured, and inundated policymakers with competing research, syntheses, and policy briefs. For instance, a 2007 Google search for “merit pay research” yielded more than 1.9 million hits. Policy-relevant research can be widely circulated and posted on the web within days or weeks of its completion, often sidestepping the peer review that precedes publication in a refereed journal.

This development is a natural response to the slow-moving and jargon-laden culture of academic publication that inhibits researchers who seek to address contemporary debates effectively. Advocates of more rapid and open dissemination argue that it has resulted in a more heterodox and timely body of education research and question whether research quality has been compromised by this shift.

To wend their way through competing analyses, journalists and policymakers rely on proxies to evaluate often-conflicting findings reported by diverse institutions operating with divergent norms. Reporters are frequently more comfortable highlighting work that draws upon federal data because they feel confident about its

provenance—though government data does not guarantee quality or neutrality. The result is that journalists and public officials may place undue emphasis on proxies for neutrality or rigor while failing to appreciate fully the importance of technical considerations like sample construction, measurement error, or internal validity.

While there are real benefits to the democratization of research, there are also substantial costs to conducting debates about the merits of research findings in public spaces. When research gets caught in larger political debates and is wielded by interested parties, technical, value-neutral arguments about sample size or measurement error can make it difficult for scholars to argue methodological questions as researchers rather than as partisans. It may be beneficial to hash out these issues within the research community rather than in press releases or newspaper stories.

Ultimately, the education community should seek to facilitate disciplined technical debate, encourage researchers to police the quality of their work, and help consumers make sense of the enormous variation in credibility of published research. This might include devising ways to support professional organizations or associations that encourage and reward self-policing; funding federally sponsored reviews of available research, as is being attempted by the U.S. Department of Education’s “What Works Clearinghouse”; increasing foundations’ attention to these considerations in the research that they fund; and fostering efforts by credible, competing institutions to assess the rigor of high-profile scholarship.

## The Role of Intermediaries

The cluttered informational environment requires that someone distill, explain, promote, and convey research to public officials if it is to be influential. While some scholars occasionally undertake these tasks themselves, most are understandably reticent to do so because academe does not reward such behavior. Consequently, the job often falls to the sprawling menagerie of intermediary organizations. The hope is that these intermediaries are conscious of quality, but incentives to be so vary significantly across organizations.

Intermediaries generally fall into one of three categories. The first category is that of expert, nonpartisan groups, such as Education Commission of the States, Editorial Projects in Education, or regional education research and development laboratories. Trading on their credibility and perceived impartiality, they have cause to

focus more heavily on synthesizing available scholarship than on actively promoting findings.

A second category includes membership groups, such as the National Education Association, Council of the Great City Schools, or the National School Boards Association. These have a responsibility and strong incentive to promote research findings that align with the interests of their members and policy agendas.

A third category includes mission-driven or ideological organizations like the Education Trust, the Heritage Foundation, the Center for Education Reform, and the Center for American Progress (CAP), which promote work that advances their ideological or philosophical approaches to school improvement.

---

NCLB has encouraged state and district officials to seek quick fixes and immediate solutions. Yet this short-term perspective is at odds with the scientific process.

---

There is a strong preference among the cognoscenti for intermediaries perceived as nonpartisan and expert, but precisely because these organizations are nonpartisan, their arguments frequently lack crisp definition. Moreover, while membership groups like the national teachers unions are able to tout research directly to major media outlets and to a network of allies in state legislatures or on Capitol Hill, nonpartisan groups do not have the same clout because they lack clear audiences of members or sympathizers.

Mission-driven groups tend to garner the most skepticism within the research community, yet these entities, like the Education Trust and the Thomas B. Fordham Foundation, nonetheless have proven immensely influential because their policy focus and energy make them effective voices. While leaders of membership groups must take care to stay in step with their members, mission-driven organizations have much more freedom and are positioned to offer clear, actionable interpretations of available research.

Intermediaries and advocacy groups inhabit an influential but murky space. Research seen as useful to the agenda of such a group can win a researcher visibility, professional contacts, access, and funding—while research that serves the interest of no organized

constituency is likely to attract less notice and yield fewer professional rewards.

The implications of this dynamic are not well understood. In practice, a researcher whose work is embraced by teachers unions, advocates for early childhood education, or the charter school community has incentives to depict his work in ways that these organizations find congenial and to remain quiescent as they apply the findings or recommendations to dissimilar circumstances. Moreover, researchers may face informal pressures from funders and allies to soft-pedal their findings if later work points to different policy implications.

For interested reformers, one intriguing response is to alter the mix of intermediaries by “stocking the pond.” Two ambitious efforts to do this in the past decade include the launch of the Washington, D.C.–based groups Education Sector and CAP. Education Sector seeks to position itself as a “neutral” voice in education, while CAP is engaged in the full panoply of policy issues and is unapologetically aligned with the Democratic Party. Each has sought to influence education debates by commissioning new research, promoting select findings, and reaching public officials.

## The Perils of Overemphasizing “Relevance”

Ultimately, the answers that research can produce may matter less than the questions and insights that it generates. Many lines of inquiry are more valuable for the questions and cautions they pose than for their ability to deliver prescriptive guidance to policymakers. One consequence of the standards and accountability movement has been to emphasize applied evaluation. While this has brought a discipline and focus to education research that was previously lacking, it has imposed real costs. The 2014 target for “universal proficiency” enshrined in NCLB and the law’s push for rapid efforts to close achievement gaps have been particularly significant here. As intended, NCLB has encouraged state and district officials to seek quick fixes and immediate solutions.

Yet this short-term perspective is at odds with the scientific process. Focusing attention and resources primarily on what might be relevant in the near term has the potential to distort research agendas and weaken support for long-term efforts to collect broad descriptive data. The research that has transformed medicine, for instance, has typically not been a product of field trials testing new medications that will be available within the decade but the searching inquiry that may take a generation to bear

fruit. An essential role for the federal government is collecting large data sets, both descriptive and longitudinal—an exercise that has lost favor of late. These efforts, housed at the National Center for Education Statistics, do not offer the immediate payoffs of more narrowly pitched evaluative work, but they are essential to sustaining vibrant inquiry.

While there is immutable tension between the desire for applicable and immediate lessons and the investigation of fundamental and long-term questions, both education research and policy would be better served if we sought a sustainable balance between the two instead of cartwheeling from one extreme to the other.

## The Political Economy of Education Research

Finally, it is vital to recognize how the “political economy” of education research helps determine what studies are conducted and how scholars approach them.

Scholars compete fiercely for the right to evaluate high-profile reform initiatives and typically require support from interested funders and access to the schools, districts, or programs under study. Winning that access is a delicate process that requires careful attention to building relationships and cultivating a reputation for probity and rigor. Whatever the intent, positive findings can yield a symbiotic relationship that serves both subject and researcher. Evaluators are inevitably more likely to study states, districts, schools, or programs where they have established cordial relationships (especially since researchers are often proponents of the reforms they are asked to examine). As a result, they have an incentive to protect those relationships and avoid being too negative when examining projects.

Practitioners and reformers are invested in their programs and are naturally predisposed to regard them as effective. These educators are frequently correct: pilot programs often impress because of the advantages of impassioned leadership, extra resources, exceptional faculty, and a common culture forged by a shared bond. They are primarily interested in evaluative research that can document the successes of their efforts and thereby open the door to additional resources and opportunities. Consequently, educators have strong incentives to provide access and data to friendly researchers, rather than those perceived as skeptical or nonevaluative.

The incentive for econometricians to work with existing data sets encourages them to study the questions

for which good data already exist and shy away from murky questions for which they do not. In this way, research and data collection on a given topic tend to attract imitators, while more difficult-to-capture questions go unexplored. In recent years, the magnetic pull of available data has driven enormous attention to measuring school performance in reading and math for children in grades three through eight and to assessing high school reform in terms of graduation rates; scant attention has been given to questions where systematic data are less readily available. At the same time, pressing policy questions such as how school districts respond to choice-based interventions, how principals respond to NCLB-style accountability, or how districts hire and assign teachers to schools have attracted little disciplined scrutiny. Federal agencies have a vital role to play in the collection of appropriately heterogeneous data.

Perhaps most significantly, educational leaders have trouble framing tractable questions or communicating clearly to researchers the kind of queries that would benefit practice, so research is frequently driven by the enthusiasm of researchers, foundations, or isolated officials in ways that hinder its ability to inform decision-making.

The result of these pressures is that the research that gets pursued is not necessarily the research that is most significant, valuable, or useful—but the research that scholars have the ability and incentives to produce. In the long run, altering the mix of research requires tackling those resources and incentives.

## Final Thoughts

Research has a vital role to play in democratic policy debate. That role is not to dictate outcomes but to ensure that public decision-making is informed by all the facts, insights, and analyses that the tools of science can provide. Researchers can challenge conventional wisdom and casual assumptions, provide realistic estimates about what reforms may or may not accomplish, help innovative ideas gain a foothold by offering credible evidence of their plausibility, and provide insight into the relative merits of multiple interventions. It is not simply a question of getting the research right. The soft tissue involved in marrying research to policy matters as much as the technical merits of research.

For instance, the federal program Reading First drew on a wealth of rigorous and sophisticated research, built upon a consensus report issued by the National Reading Panel, and made an unprecedented federal investment in

reading with strong bipartisan backing. Nevertheless, the program's awkward construction and predictably problematic implementation compromised its natural advantages. The political and legal travails of Reading First have raised questions about the program's legitimacy, undermined support for its funding, and illustrated how perilous this course can be if it is not informed by attention to institutional design and political dynamics.

The dispassionate "scientist" is an uncommon figure in education policy debates; moreover, it is not clear that he should be the ideal. It would be strange indeed, as Stanford University professor Terry Moe has eloquently observed, if scholars did not have strong, informed opinions about subjects that they have spent years or even decades studying.<sup>12</sup> It would be a peculiar kind of reticence that encourages a scholar to remain mum about her own conclusions and sit on the sidelines even as others—less expert—opine freely, using (and misusing) her work.

Nonetheless, if academics are to fulfill the valuable social function of serving as independent sources of insight and knowledge, they must retain the ability to champion policies and ideas, ask hard questions, and change their minds while remaining somewhat removed from the partisan contests and political projects of the moment. As relevance in policy debates requires alliances with the intermediaries and policymakers who make things happen, this is inevitably a delicate dance.

This state of affairs encourages successful researchers to adopt one of two courses: either focus on narrow technical work and studiously avoid offering opinions on policy or become enmeshed with one side or another in heated public debates. Neither course seems optimal, but the resources, professional norms, and incentives that might encourage researchers to negotiate a middle path are in short supply. How the research community wrestles with this tension in the years ahead will be critical to the nature and influence of education policy research.

Finally, there are steps that researchers, the federal government, foundations, and the profession's leadership can take that will benefit researchers who are careful about respecting data and avoiding careless claims and that can incline researchers to seek a healthy independence from partisan conflict. These involve helping policymakers and the public understand what research can and cannot contribute, supporting self-policing within the research community, encouraging the development of professional norms about what constitutes appropriate involvement in

public debate, steering more funding toward research that is vetted by knowledgeable researchers, and investing more heavily in large public data sets.

How best to accomplish these goals without stifling far-reaching inquiry or unduly narrowing scholarship is a question that researchers and policymakers must wrestle with in the years to come. In the end, the tension between those engaged in accumulating knowledge and in making policy is a frustrating but essential one in a democratic nation.

---

*AEI web editor Laura Drinkwine worked with Mr. Hess to edit and produce this Education Outlook.*

## Notes

1. Douglas N. Harris, "Class Size and School Size: Taking the Trade-Offs Seriously," in *Brookings Papers on Education Policy: 2006/2007*, ed. Tom Loveless and Frederick M. Hess (Washington, DC: Brookings Institution Press, 2007), 137–61, available at [www.aei.org/book903/](http://www.aei.org/book903/).

2. American Educational Research Association, "Class Size: Counting Students Can Count," *Research Points* 1, no. 2 (2003): 4.

3. Diane Whitmore Schanzenbach, "What Have Researchers Learned from Project STAR?" in *Brookings Papers on Education Policy: 2006/2007*, 206.

4. *Ibid.*, 213.

5. Peter Schrag, "Policy from the Hip: Class-Size Reduction in California," in *Brookings Papers on Education Policy: 2006/2007*, 235.

6. CSR Research Consortium, "What We Have Learned About Class Size Reduction in California," September 2002, available at [www.classize.org/techreport/CSRYear4\\_final.pdf](http://www.classize.org/techreport/CSRYear4_final.pdf) (accessed January 3, 2008).

7. *Ibid.*, 4–5.

8. Peter Schrag, "Policy from the Hip: Class-Size Reduction in California," 235–36.

9. Eric Hanushek, "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis* 21, no. 2 (1999): 143–63.

10. Gerald Horton, "Guest Editor's Introduction," *Social Research* 72, no. 1 (Spring 2005): viii, available at [www.socres.org/vol72/721intro.pdf](http://www.socres.org/vol72/721intro.pdf) (accessed January 18, 2008).

11. Quoted in Gerald Horton, "Guest Editor's Introduction," viii.

12. Terry M. Moe, *Schools, Vouchers, and the American Public* (Washington, DC: Brookings Institution Press, 2001).