



## Science and Nonscience: The Limits of Scientific Research

By Frederick M. Hess

*Assessing many educational practices through scientific principles can be useful, but reformers must ensure that the push for scientific inquiry does not stifle reforms, such as promoting flexibility, competition, and accountability, that do not lend themselves as readily to rigorous scientific evaluation.*

American education research has turned a corner. The 2002 creation of the Institute of Education Sciences (IES), the ascendance of accountability, and the No Child Left Behind Act's demand for "scientifically based research" have radically altered an educational research culture that just a few years ago bridled at the "medical model" and too often championed ethnographies, action research, "critical narrative," "discourse analysis," and other approaches that provided parents, practitioners, and policymakers with little useful information.

Together, both NCLB and IES represent a demand that rigorous scientific principles be used to assess programs. This development did not "happen," and it was not an inevitable evolution embraced by the education research community. Rather, this change was the consequence of prodigious efforts by proponents like Congressman Michael Castle (R-Del.), reading expert Reid Lyon, and IES head Russ Whitehurst. For their efforts, they have met with fierce resistance from some quarters of the education research community, as well as professional discourtesy, bizarre conspiracy theories, and ad hominem attacks.

---

Frederick M. Hess ([rhess@aei.org](mailto:rhess@aei.org)) is a resident scholar and the director of education policy studies at AEI. His most recent book is *Urban School Reform: Lessons from San Diego* (2005). A version of this article appeared in the *Education Gadget* on February 17, 2005.

The notion that education ought to hold science in the same high regard as do medicine and engineering would seem axiomatic. In principle, IES's mission to transform education "into an evidence-based field in which decision makers routinely seek out the best available research and data before adopting programs or practices" is entirely to the good. The changes have focused researchers on questions of validity, reliability, and replicability, and have raised the bar for the investment of federal funds.

Amidst this good news, however, lurks the risk that the pendulum will swing too far, that the lure of "scientifically based research" will cause certain methods of study—especially randomized field trials—to be demanded even when ill-suited for the issue at hand. This risk looms in light of the new research "priority" published in the *Federal Register* on January 25 that dramatically expands the scope of activities subjected to "scientifically based research."

### Judging Reforms Scientifically

I am—emphatically—not issuing here another plea for "mixed methods" nor expressing concern about the practice of science. Instead, I am raising a more concrete and practical concern: that we risk stifling sensible and promising structural reforms in schooling. This risk is posed when we

start to imagine that reforms to personnel, management, or financial systems need to be subjected to these scientific standards. In such cases, a premature or unyielding application of the tenets of “scientific research” could insulate ineffective and dysfunctional arrangements from needed and attainable reform.

How does this danger arise? In large part, it occurs from an imperfect understanding of how the “medical research model” works in medicine and how and when to import it into education. It is vital to recognize that there are really two kinds of “reforms” in medicine or education—and that the proper role of science and scientifically based research is very different from one to the other. One kind of reform relates to specialized knowledge of how the mind or body works, and the other relates to the manner in which we design and operate organizations, governments, and social institutions.

In education, the former category deals with the science of learning and with behaviors and programs that induce it. Such measures include pedagogical and curricular practices and interventions that relate to the development, knowledge, skills, and mastery of individual students. Relevant approaches would include methods of literacy instruction, bilingual education, sequencing mathematical subjects, and so on. Each of these entails the application of discrete treatments to identifiable subjects under specified conditions in order to achieve specific ends. Such interventions are readily susceptible to field trials, and findings on effectiveness can reasonably be extrapolated to other populations. It is desirable and appropriate that such reforms be subjected to rigorous empirical evaluation (and, whenever possible, to randomized field trials) and that educators be encouraged, even pressed, to use demonstrably superior approaches—and to eschew those lacking such evidence.

The second category of reform entails governance, management, or policy innovations intended to improve organizational effectiveness. It includes such innovations as permitting mayors to appoint school boards, permitting schools to operate free of some regulations, paying employees based on performance, and so on. None of these changes is unique to education. They draw upon a

mass of experience gained in other sectors—and their effects are consistent enough and understood well enough across a broad swath of human experience that it is neither useful nor appropriate to use the scientific method to determine whether, for example, initiatives to reward excellence, increase managerial flexibility, or

---

Reforms that address pedagogy, curricula, or teaching practices are fundamentally different from those that seek to change the incentives under which educators operate, how much flexibility they have, or how they are hired, managed, or held accountable—and should be treated as such.

---

ensure accountability may hold promise in schooling. Such interventions are rarely precise, do not take place in controlled circumstances, and generally are administered to classes of people rather than discrete clients. Since the results of these structural reforms will be contingent on the context and manner in which they are implemented, even well-designed studies will find it problematic to draw lessons from isolated experiments that trump our broader body of knowledge regarding the use of incentives or markets. Of course, we should welcome inquiry and take new findings into account when reflecting on policy or program design. However, it is vital to remember that we have got a vast store of knowledge on these questions, and that whatever the results of small-scale experiments with merit pay or educational competition, this existing body of knowledge ought to weigh more heavily than the results of one or another context-specific study.

## The Limits of Educational Science

Reforms that address pedagogy, curricula, or teaching practices are fundamentally different from those that seek to change the incentives under which educators operate, how much flexibility they have, or how they are hired, managed, or held accountable—and should be treated as such. For instance, in medicine, while we deem it appropriate for the Food and Drug Administration to monitor and approve drug therapies and treatments, we do not require FDA approval before we permit doctors, hospitals, or health care firms to change their management practices, compensation strategies, accountability metrics, or work routines.

In truth, charter schooling, accountability systems, school vouchers, alternative certification, and merit pay are not really “educational” innovations in any meaningful sense. They do not rest on conceptions of teaching or

learning processes or practices in the way that decisions about literacy or math programs do. They are decisions about how to arrange and deliver services, similar to those made in social welfare, library management, higher education, or private enterprise. Such decisions draw upon our experience across a wide range of human endeavors and organizations. They apply practical wisdom and experience about human behavior from a wealth of sectors. We should welcome research on the effects and efficacy of such reforms and use them in debating and crafting policy. But we also need to understand the limits of science.

The notion that rewarding performance ought to be subject to scientific validation before adoption is akin to suggesting that the National Institutes of Health should determine permissible compensation systems for doctors.

If we applied that logic to other agencies of state government, we may well never have automated state revenue departments, streamlined departments of motor vehicles, or permitted states to reward whistleblowers who expose fraudulent activities.

As we seek to build a scientific knowledge base in education, after a century of dawdling, we should be careful not to swing the pendulum so far that we come to regret it. While the elevation of “educational science” is laudable, it is important to keep it in perspective. The push for scientific inquiry should not undermine sensible efforts to promote flexibility, competition, efficiency, and accountability. Those who want school reform to track both science and common sense must take care that proper respect for science is accompanied by a similar respect for the limits of science.