

EDUCATION WEEK

Published Online: August 25, 2009

Published in Print: August 26, 2009

Commentary

Innovative Reforms Require Innovative Scorekeeping

By Lisbeth B. Schorr

President Barack Obama has made clear that we must systematically identify “what works,” both for budgetary reasons and to ensure that public money supports effective social programs and policies. The president and his budget chief recognize how tricky it is to make that determination. In June, Office of Management and Budget Director Peter Orszag **released a statement** describing how the administration will make sure that spending decisions are based not only on good intentions but also on strong evidence.

Serious social reformers today agree that rigorous efforts to determine “what works” are essential. But, depending on what the administration considers “strong evidence,” these efforts risk sabotaging or marginalizing some of the most innovative attempts to solve intractable social problems. I worry that, in defining what constitutes “the best available

No single, circumscribed program can turn things around in an entire community or for a whole population. Nor can complex social programs and policies be tested like new drugs.

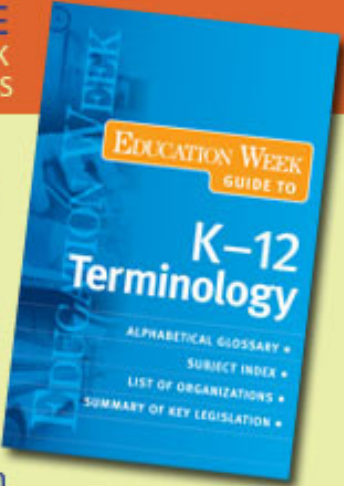
evidence” of effectiveness, the OMB and federal agencies will follow the constricted approach of the Coalition for Evidence-Based Policy and the U.S. Department of Education’s What Works Clearinghouse. These and similar organizations claim scientific rigor by insisting that public and philanthropic support go only to

[← Back to Story](#)

NOW AVAILABLE
FROM EDUCATION WEEK
PRESS AND JOSSEY-BASS

“This book cuts through the jargon, and provides clear definitions written in plain English.”
Anne L. Bryant
executive director,
National School Boards
Association

ORDER TODAY
www.josseybass.com



programs shown to be evidence-based through *experimental evaluation methods*, preferably involving random assignment of participants to experimental and control groups. The implication is that this methodology can determine definitively and objectively, uncontaminated by human judgment, whether any intervention—be it a pill, a model program, or an ambitious institutional change—produces a different outcome than would otherwise occur.

Unfortunately, no single, circumscribed program can turn things around in an entire community or for a whole population. Nor can complex social programs and policies be tested like new drugs. The interventions that turn around inner-city schools, strengthen families, and rebuild neighborhoods are not stable chemicals manufactured and administered in standardized doses. They are sprawling efforts with multiple components, some of which may be proven experimentally, but many that can't be because they require midcourse corrections and adaptations to fit local circumstances.

Reformers in virtually every domain—from education to human services and social policy—have been learning that the most promising strategies are likely to be complex and highly dependent on their social, physical, and policy context. Very few efforts to improve education for at-risk students, prevent child abuse, increase labor-market participation, or reduce teenage pregnancy or homelessness succeed by applying a single, bounded intervention. They depend on community capacity to take elements that have worked somewhere already, adapt them, and reconfigure them with other strategies emerging from research, experience, and theory to make a coherent whole.

The search for silver bullets is giving way to an understanding that, to make inroads on big social problems, reformers must mobilize multiple, interacting strategies that take account not only of individual needs but also of the power of context. President Obama **has urged that** we stop treating unemployment, violence, failing schools, and broken homes in isolation and put together what works “to heal that entire community.” That’s the thinking behind the president’s proposed Promise Neighborhoods initiative, inspired by the accomplishments of the **Harlem Children’s Zone**.

What is remarkable about the collection of activities that the Harlem program comprises, and what has captured the attention of funders, reformers, and politicians, is that they build on one another; each is shaped to add to and multiply the impact of the others. Theory and experience suggest that the long-term results of these coherent efforts will ultimately be a critical mass of engaged, nurturing families, well-educated students, community values that support education and responsibility, and an infrastructure to sustain results

that cannot be achieved by isolated programs aimed only at individuals.

The trouble is that scaling up such collections of reforms is hard, and determining what, exactly, works is even harder.

In assessing the success of complex, interactive efforts to improve outcomes, experimental methods cannot be the sole arbiter of effectiveness.

As a family-support program in King County, Wash., has discovered, the “rigid, narrow accountability” that funders demand forces programs to “keep doing only what worked yesterday, instead of what works today.”

In an internal evaluation, it found that the very qualities that make the program effective are the qualities that make measurement so difficult.

The obstacles to demonstrating effectiveness, which become even more formidable in moving beyond the programmatic, are best overcome with a clear focus on results.

In the 1990s, the state of Vermont established state-local partnerships so people in all domains could do everything likely to contribute to school readiness. Their focus on results encouraged innovation and local problem-solving and replaced rigid regulation of inputs with rigorous accountability for accomplishments.

Vermont leaders knew they would never be able to *prove* that each piece of what the partnerships did was effective, but they were able to show that the entire strategy dramatically improved lives. Trend lines that had shown increasing damage in the form of child abuse, infant mortality, school failure, and teenage pregnancy began to turn around and move in the right direction soon after the partnerships instituted policies targeting those outcomes.

The evidence came from timing (the curves began to turn in communities where the interventions were initially implemented, and then in the whole state as the interventions went statewide); from theoretical connections established by research (for example, that high-quality supports to young families can reduce child abuse and changed community norms can reduce teenage pregnancy); and from the accumulation of data (including practitioner observations and official data from hospitals, health departments, and schools).

Had the Vermont partnerships been limited to “proven” interventions, or had they tried to set up interventions as randomized experiments, they would have had neither the money nor the flexibility to provide the services that made such a remarkable difference for the state’s children and families.

When an orientation toward results pervades planning, management, and implementation of new initiatives, it is easier to meet the challenges of accountability and evaluation. Evaluation becomes a way to support rigorous, contemporaneous collection of data on progress toward clearly defined goals, rather than an after-the-fact assessment of what succeeded (or didn't).

Developers of complex social reforms aren't the only ones who find that experimental methods are not always the best or even most "scientific" way to obtain credible evidence. Calls to re-examine what constitutes credible evidence come even from medicine. The Roundtable on Evidence-Based Medicine of the federal Institute of Medicine recommends that randomized clinical trials should *not* continue to be considered the gold standard, as they seem useful only in limited circumstances, including a narrow range of illnesses and the absence of multiple problems in an individual patient.

Experimental methods are not always the best of even most 'scientific' way to obtain credible evidence.

Many education researchers have reached a similar conclusion. In the American Educational Research Association's *Handbook of Education Policy Research*, David L. Weimer suggests that "the typical evaluation model focuses attention on one or a small number of policy impacts with unambiguous desirability, and only assesses policies already in place." He points out that truly novel ideas cannot be assessed within this model because they have yet to produce data that can be used to measure impacts.

Policymakers radically diminish the potential of reforms if they allow themselves to be bullied into accepting impoverished definitions of credible evidence. Just as the Obama administration is on the cutting edge of reform by recognizing the importance of complexity in many arenas of social policy, so must it encourage innovation in efforts to determine "what works."

Lisbeth B. Schorr is a senior fellow at the Center for the Study of Social Policy, in Washington, and a lecturer in social medicine at the Harvard Medical School, in Boston. She is the author of Within Our Reach and Common Purpose.