

Conference of the National Alliance of Multi-ethnic Behavioral Health Associations
Keynote
Lisbeth B. Schorr¹
Integrating Evidence and Practice to Reduce Disparities:
Developing an Inclusive Framework for Effective Mental Health Interventions
May 11, 2010

Ever since Mareasa Isaacs invited me to speak here today, I've been looking forward to meeting with this group of people to discuss how we might think about evidence in ways that would strengthen our capacity to solve persistent social problems and reduce disparities in outcomes.

In the January 2010 issue of Child Development, Jack Shonkoff, the pediatrician founder of the Harvard Center on the Developing Child, points out that even today's evidence-based programs are not having the magnitude of effect needed to change outcomes on a large scale for the children and families most at risk. To do better for these populations and reduce disparities, we cannot simply rely on spreading what has been shown to work in the past. Rather we must analyze past successes – and failures – so that we might innovate, combine formerly separate interventions, and generate new hypotheses that will generate improved outcomes and reduced disparities in the future.

But the calls to innovate, now coming from far and wide, including from President Obama, are somehow coupled with an insistence that public and philanthropic funds should go only to what has been shown to be “evidence-based.” That is a tension that could actually interfere with efforts to reduce disparities -- **unless we agree on an inclusive definition of evidence, and view current evidence-based practices as a starting point to build on, rather than a final destination.**

It may be useful at this point to review the history of how we got to where we are today, when -- paradoxically -- our expanding knowledge base about what works exists side-by-side with new constraints on what counts as legitimate innovation and experimentation

I go back to the 1960s –

Whiz Kids came from Pentagon to the Office of Economic Opportunity

¹ Lisbeth B. Schorr is Senior Fellow at the Center for the Study of Social Policy and Lecturer in Social Medicine at Harvard University

No one paid much attention to their esoteric calculations. (Except for the 17 to 1 cost effectiveness of family planning, which got a lot of attention, it was mainly parallel play.)

The early evaluation of Head Start by Westinghouse, which found it didn't raise the IQs of participating children, had little impact in the face of Shriver's vision that even in its first summer, OEO would "write Head Start across the face of this nation so that no Congress and no President can ever destroy it."

It was an era when a compelling vision, good intentions, and heart-rending stories were enough.

As we all know, that era didn't last. The shine wore off the programs aimed at ending poverty in the face of Viet Nam and other competing demands, occasional failures, a few scandals, and – ultimately -- the Reagan and Reaganesque campaigns to shrink government. The calls for evidence of effectiveness became louder, usually in the form of demands for evidence from randomized experiments. The idea that you could actually PROVE that a social intervention worked by using "scientific" methods was thrilling. Funders – public and philanthropic – didn't have to make fallible judgments – they could rely on proof that came out of incontrovertible numbers. Economist Rob Hollister says it's "like the nectar of the gods: once you've had a taste of the pure stuff it is hard to settle for the flawed alternatives."

The prevailing wisdom held that we could evaluate social programs with the same methods that had led to the nation's great medical advances. Economist Alan Krueger urged that we test education reforms the way the Food and Drug Administration tests drugs. A Brookings Institution symposium concluded that finding out what works in educational and social programs requires randomized field trials, in which "one person gets the pill, and the other person gets the placebo." Those who had the sense that there was a mismatch here were afraid to speak up, lest they be seen as "unscientific." Both the producers and consumers of evaluation research allowed themselves to be bullied into accepting narrow and impoverished definitions of evidence as the only evidence worth having.

Daniel Yankelovich and I considered ourselves very brave when we wrote in the Los Angeles Times during the 2000 presidential campaign that evaluating complex social programs is not like testing a new drug, that promising social programs tended to be sprawling efforts with multiple components requiring constant mid-course corrections, the

active involvement of committed human beings and flexible adaptation to local circumstances. We warned that the interventions that were needed to rescue inner-city schools, strengthen families and rebuild neighborhoods were not stable chemicals, manufactured and administered in standardized doses.

Over the last decade, in one field after another, as we began to recognize the complexities of “what works” and what needed to be done, it became increasingly clear that no single approach to evaluation would allow us to learn enough from past interventions; nor would it enable us to predict the success of future efforts.

The costs of randomized trials seemed excessive, and the results took so long to arrive that by the time they did, they were often irrelevant

The experimental evaluations didn’t provide enough information about implementation and *how* the work was affecting outcomes

It seemed as though the interventions most likely to result in significant improvements in outcomes, especially among the most disadvantaged, were hardest to assess in traditional ways, because they

Were complex, interactive, and relationship-based

Had to be adapted to a variety of cultures and populations, as well as to new and changing contexts even during initial implementation

Required significant front-line flexibility and sensitivity

Were continually evolving in response to changes in context, lessons learned, and advances in knowledge

Required reforms not just at the program level, but also at the institutional, policy, systems and population level

Were unlikely to consist of isolated programs, operating within isolated silos, but needed to combine proven and promising practices in new ways; often the best practices that have to be “bundled” together had to come out of separate systems, including education, juvenile justice, health, child welfare, and early childhood.

Because we didn’t have the tools to assess these complex – and often innovative – interventions, they were sometimes simply discarded. When they were undertaken and evaluated by non-experimental and more inclusive methods, they never made the lists of “proven programs.”

It is fascinating that the first concerted thrust toward a more inclusive definition of evidence came from medicine -- which had pioneered and made the greatest advances using randomized experiments.

Don Berwick, the medical reformer who has just been nominated by President Obama to head the Centers for Medicare and Medicaid Services, wrote that the RCT is a powerful, perhaps unequaled, research design, but only to explore the efficacy of selected components of practice -- drugs, procedures, and other interventions that are conceptually neat, and have a linear, tightly coupled, causal relationship to the outcome of interest. Berwick and other medical experts are now saying that although the benefits of evidence-based medicine have been immense, “we have overshot the mark.” They find that the commitment to “evidence-based medicine *of a particular sort* into an intellectual hegemony” has created “a wall, not only in medicine but in social and educational policy, that excludes too much of the knowledge and practice that can be harvested from experience.”

A group assembled by the Institute of Medicine to inform decision-making about obesity prevention issued its report in April, calling on “the scientific community to rise to the challenge and transform the evidence picture to be commensurate with the needs of policy makers and funders.” It concluded that a more useful approach would expand evidence by crossing disciplines, making connections among population-level and community-level influences, and obtaining the practice-based evidence that is intrinsically relevant to natural settings.

Similar views have been expressed in the domains of early childhood, education, child abuse and neglect, and international development.

The Government Accountability Office has also weighed in, warning against OMB’s elevation of experimental evidence to the top of the methodological hierarchy. In a report released last November, the GAO concluded that requiring evidence from randomized studies as the sole proof of effectiveness will likely exclude many potentially effective and worthwhile practices.

What we have learned, then, in the past two decades about what makes for effective interventions, together with an increasing openness to multiple ways of knowing, puts us in a position to be more intentional about how we figure out what to do and how to do it, and enables us to assemble a much richer knowledge base.

Current experience around the country suggests that this process is best begun with a focus on results. When communities or collaboratives agree on outcomes and results for children and families, they can encourage innovation and local problem-solving that responds to unique conditions and cultures, by replacing rigid regulation of inputs with rigorous accountability for accomplishments. A results frame makes it easier to bridge diverse constituencies and points of view, support collaboration across boundaries, mobilize and sustain efforts over time. They can select the actions they will undertake by mapping backwards from the agreed-upon results, and determining what collection of interventions and supports are most likely to achieve the results.

Let me give you a few examples of how we're in a much better position today than we were a couple of decades ago, to gather evidence about what works, including practice-based evidence, and to make our decisions about what to scale up and spread much more evidence-informed.

I must warn you, however, that – while some of my best friends are in mental health – several of the examples I'm going to cite have only a tangential connection to mental health interventions, but I'm betting that you will be able to find very similar examples from your own experience.

First, **we have a much better sense of what to measure.** From a systematic analysis of practice, we know a lot more than we used to about what really matters.

Most of you will be too young to recall the famous 1960s Coleman study. It got a lot of attention because it was the largest social science research project undertaken to that time. The distinguished group of researchers announced their conclusions with considerable fanfare: it wasn't schools but family background that determined student achievement. Today we know that schools are able to change outcomes among children from profoundly disadvantaged family backgrounds. We also know where the Coleman team went wrong. It was the old evaluation mindset that led them to take into account only the most readily quantifiable data, such as per-pupil expenditures and the number of books in school libraries, components of schooling that would turn out -- years later -- not to be critical factors in changing outcomes. Coleman's team had paid no attention to what practice has shown really matters, including teacher quality and school climate. No wonder schools were found to be irrelevant to outcomes.

Second, **we are able to go deeper than yes/no, it works or it doesn't. We can determine what aspects of implementation are crucial.**

When the first large national study of Big Brother's Big Sisters came out, it seemed as though the mentoring that so many observers and advocates had bet on, wasn't changing lives after all. It was only after the evaluators unpacked their findings to look at the quality of the relationships, and at the quality of the recruitment, training, matching, and supervision, that it was possible to identify the patterns of implementation that led to improved outcomes in some of the sites, and that later became core principles.

Across domains, the growing field of implementation research has identified generic components of effective implementation, which include intensive attention to *staff selection, preservice and inservice training, ongoing consultation and coaching, staff and program evaluation, and facilitative administrative and systems support*. These interactive processes are integrated to maximize their influence on the organizational culture.

Third, we no longer have to get stuck assessing only individual programs, but can **extract the commonalities among various models**, and ask the implementers to adhere to those fundamental components with fidelity, rather than to attempt to clone individual models.

Last Fall, executives from the Cleveland Clinic, Geisinger Health System, Guthrie Health, Billings Clinic and seven other health care systems that have been cited by President Obama, Congressional leaders and policy experts as medical exemplars met under the aegis of the Commonwealth Fund and the Rockefeller Foundation in WDC. Although each organization is different, the group identified important similarities: doctors are paid a salary and patients are generally given care coordinated among a variety of specialists; leadership by a doctor, use of electronic medical records and the requirement that doctors cooperate across specialties. These are the principles they suggested for incorporation into health reform legislation.

Fourth, we now have new and **better ways to use our knowledge base -- by identifying the best practices that we can build on and combine with other evidence-informed practices**. Especially if we are to reach the children and families most at risk, we cannot keep doing only what worked yesterday, instead of building on yesterday's lessons to implement what is most likely to work better tomorrow.

Eg. Nurse Family Partnership – not successful with depressed moms, don't enroll moms involved with substance abuse

If you wanted to link up with services for these populations, your home visiting program would look different in every community, and couldn't be as elegantly evaluated as the Nurse Family Partnership has been.

Lastly, we've learned about the importance of context and infrastructure. If single, isolated remedies rarely improve results among the most vulnerable precisely because they change only one thing at a time, we need an infrastructure that holds these isolated programs together, makes them coherent, monitors quality, and sustains them. This is as true for the systems of care that you work with, as it is for the incipient Promise Neighborhoods I've been working with. Both are vastly more than the sum of their parts.

Conclusion

If we were to systematically accumulate our findings from theory, experience and research, across silos and across funders, each individual study wouldn't have to bear the full burden of methodological rigor to inform future actions and decisions. We could zero in on the commonalities among efforts that achieve the outcomes we seek, recognizing that complexity is inherent in both the problems and solutions we are most interested in.

The new approaches to knowledge building do not reject -- on grounds of messiness or complexity -- information that can shed light on real-world efforts that promise to improve outcomes. They insist on rigor even in the absence of certainty or proof, in the belief that credible evidence of effectiveness can be attached to promising interventions that do not lend themselves to experimental evaluations if they are found to have at least

- A basis in strong theory
- A converging accumulation of empirical evidence of effectiveness from similar or related efforts that may not rise to the level of causal proof.
- Consensus among informed experts based on a combination of theory, research, and practice-based evidence.

In conclusion, I want to thank you for inviting me to meet with you on this enormously important subject. As you pursue these issues over the next two days, as you deliberate on what it will take to do your work at the highest levels of quality and effectiveness, know that you are not alone in struggling with the challenge of learning in all the ways we can, from practice and theory and research, how we can improve outcomes and reduce disparities among the vulnerable children and families you serve.