



AN EVIDENCE FRAMEWORK TO IMPROVE RESULTS

Background Paper for the
**2014 HAROLD RICHMAN PUBLIC POLICY SYMPOSIUM:
THE FUTURE OF EVIDENCE**

Washington, DC
November 13, 2014

by
**Lisbeth Schorr and Frank Farrow
with Joshua Sparrow**

The Friends of Evidence
at
The Center for the Study of Social Policy

Center
for the
Study
of
Social
Policy

ABOUT THE FRIENDS OF EVIDENCE

*The **Friends of Evidence** came together quite informally, beginning in 2011, under the impetus of Frank Farrow, Director, Center for the Study of Social Policy, Lisbeth B. Schorr, Senior Fellow, Center for the Study of Social Policy, and Joshua Sparrow, MD, Director of Strategy, Brazelton Touchpoints Center, Boston Children's Hospital. The dozen of us who made up the **Friends of Evidence** from the outset had in common a passionate interest in improving outcomes for the children and families who were not faring well in today's society, though we came at these challenges from diverse perspectives.*

As we engaged each other in conversation and interviewed some of our colleagues to obtain a wider set of views, we found that our varied backgrounds and experiences had led us to shared observations and convictions about the role of evidence in efforts (public and philanthropic, local, regional, and national) to improve outcomes and to ensure the wise allocation of scarce resources. We also found that a growing number of distinguished leaders, across diverse fields, disciplines, and sectors, were drawing similar conclusions about the limitations of the recent evidence mindset, and were seeking each other out to define a more effective one. We were eager to explore these issues together and agreed to meet periodically and to seek support, hoping that our explorations would lead us to some critical insights that we might sharpen and act on with others.

We organized the 2014 Symposium on The Future of Evidence as part of our efforts to refine our understandings of the role of evidence in all our work, and to expand the circle of concerned colleagues who share our interest in becoming more strategic—in our thinking as well as in our actions—about how to generate, analyze, and apply evidence to improve critical societal outcomes and to ensure the wise allocation of scarce resources.

We thank the Annie E. Casey Foundation and the Ford Foundation for their support of the 2014 Symposium and the work that will grow from it.

THE FRIENDS OF EVIDENCE

Susan Bales

President, FrameWorks Institute

Anthony Bryk

President, Carnegie Foundation for the Advancement of Teaching

Deborah Daro

Senior Research Fellow
Chapin Hall Center for Children, University of Chicago

Frank Farrow

Director, Center for the Study of Social Policy

Lawrence Green, DPH

Professor of Epidemiology and BioStatistics
UCSF School of Medicine

John Kania

Managing Director, FSG

Nat Kendall-Taylor

Research Director, FrameWorks Institute

Patti Patrizi

Patrizi and Associates

Charles Payne

Professor, School of Social Service Administration
University of Chicago

Lisbeth B. Schorr

Senior Fellow, Center for the Study of Social Policy

Joshua Sparrow, MD

Brazelton Touchpoints Center
Boston Children's Hospital and Harvard Medical School

CONTENTS

EXECUTIVE SUMMARY	i
I. CHANGING TIMES REQUIRE MORE POWERFUL RESPONSES	1
WE ARE TACKLING TOUGHER PROBLEMS.....	2
EARLIER RESULTS HAVE BEEN LIMITED.....	2
PREVAILING WAYS OF OBTAINING EVIDENCE ARE MISMATCHED TO NEW WAYS OF WORKING.....	3
II. EARLY RESPONSES	5
CALLS FOR GREATER USE OF EVIDENCE.....	5
DEVELOPMENT OF FASTER, CHEAPER RANDOMIZED TRIALS.....	6
APPLYING TIERED HIERARCHIES OF EVIDENCE TO FUNDING DECISIONS.....	7
III. BUILDING A STURDIER EVIDENCE FRAMEWORK	8
1. SYSTEMATIC LEARNING IS THE FOUNDATION.....	9
2. ACCOUNTABILITY FOR RESULTS IS AN ESSENTIAL PILLAR.....	12
3. USEFUL AND USABLE MEASURES ARE MADE MORE READILY AVAILABLE.....	13
4. COMPLEXITY IS FULLY RECOGNIZED.....	14
5. A FULL RANGE OF EVIDENCE IS MOBILIZED IN DECISION-MAKING.....	16
IV. OPERATIONALIZING A POWERFUL APPROACH TO EVIDENCE	20
ELEMENT 1: MANY SOURCES OF EVIDENCE INFORM INTERVENTION DESIGN.....	22
ELEMENT 2: RESULTS SHAPE IMPLEMENTATION.....	25
ELEMENT 3: NETWORKS ACCELERATE KNOWLEDGE DEVELOPMENT AND DISSEMINATION.....	31
ELEMENT 4: MULTIPLE EVALUATION METHODS FIT DIVERSE PURPOSES.....	32
ELEMENT 5: A STRONG INFRASTRUCTURE SUPPORTS CONTINUOUS LEARNING FOR IMPROVEMENT.....	35
V. NEXT STEPS	36

EXECUTIVE SUMMARY

Given that we as a nation pride ourselves on our pragmatism and ingenuity, our efforts to become more effective in addressing today's most serious social problems require the most inclusive approach to learning about what works to improve results and to assure the responsible expenditure of public and philanthropic funds. In the face of the many obstacles that stand in the way of the progress we all want to see, we suggest that **we can overcome one significant obstacle by expanding our understanding of what constitutes useful and usable evidence.**

The question we address in this paper is not whether we need evidence, or even whether we need more evidence. **The question we explore here is, How do we use a framework of continuous learning to obtain and apply the kinds of evidence that will be most useful in achieving significantly greater outcomes?**

Changing Times Require More Powerful Responses

Many of our current structures for generating and using evidence to improve outcomes were developed to assess programmatic interventions that represented the frontier for their time. Today, as we tackle tougher problems with more complex interventions, we see that many of our new ways of working cannot be understood or improved using only a limited range of evidence tools.

Early Responses

We identify several efforts now underway to respond to changing circumstances: the call for greater use of evidence; support for developing faster, cheaper randomized trials; and the application of tiered hierarchies of evidence to funding decisions. These steps toward more, and more usable, evidence represent, in different ways, expanded approaches to how we think about evidence.

Building a Sturdier Evidence Framework

It will take a wider range of evidence-generating methods that are widely understood, endorsed, and applied to provide the timely, actionable, real-world knowledge and the climate of learning to achieve substantially better outcomes, more-effective use of resources, and greater accountability. We identify the components of a sturdier evidence framework that is designed to be a better fit with the complex interventions being implemented today:

1. Systematic learning is the foundation.
2. Accountability for results is an essential pillar.
3. Useful and usable measure are made more readily available.
4. Complexity is fully recognized.
5. A full range of evidence is mobilized in decision-making.

We see this evidence framework as a basis for producing more and better information—if not always more-certain information—that will contribute to the progress we're looking for, and that will shape a shared understanding of what evidence-based means. To be solidly evidence-based, decisions about programs, policies, strategies, and practices must be the product of multiple efforts to understand them.

Operationalizing a Powerful Approach to Evidence

Today, leaders at many levels and in diverse domains are operationalizing a more powerful approach to evidence. They take multi-faceted approaches to generating and using evidence as part of their efforts to reform systems, transform conditions in communities, and continuously improve and adapt their own interventions. We identify five elements of a more powerful approach to evidence that are currently in real-world operation.

1. Many sources of evidence inform the understanding of needs, assets, and context, and the consequent intervention design.
2. Results shape and reshape implementation.
3. Networks accelerate knowledge development and dissemination.
4. Multiple evaluation methods fit diverse purposes.
5. A strong infrastructure supports continuous learning for improvement.

We illustrate how the elements of the more powerful approach to evidence are being operationalized by the following initiatives:

- **The CareerAdvance® Program**, a work-readiness program of the Community Action Project of Tulsa, Oklahoma
- The James M. Anderson Center for Health Systems Excellence at **Cincinnati Children’s Hospital Medical Center**, which focuses on transformational quality, process and outcomes improvement
- **Community College Pathways**, a Networked Improvement Community of 27 community colleges and three major universities
- **LIFT**, which operates in six cities to help build the personal, social and financial foundations its members need to get ahead
- **The MOMS Partnership** of New Haven, Connecticut, a community-based initiative that addresses maternal depression
- **The Northside Achievement Zone**, a Promise Neighborhoods initiative in Minneapolis

None of the five elements of the more powerful approach to evidence that we describe and illustrate is new; all have been identified and examined in existing bodies of knowledge, and each is currently in use somewhere. **What has not yet happened, except in isolated instances, is the intentional and systematic creation of the conditions under which these diverse ways of generating knowledge can be coherently implemented, rigorously examined, and supported over time.** That is what we hope to see as the “new normal” of the future.

Next Steps

We are impressed by the many individuals and organizations throughout the country that are struggling with the challenge of generating, applying, and disseminating knowledge that is strong enough, deep enough and scalable enough to be a credible guide to more effective interventions and—ultimately—to substantially improved outcomes. Ideas that emerge from the symposium will form the next stage of work by the Friends of Evidence to chart the most feasible steps and, hopefully, to assist the federal government, foundations, practitioners and opinion leaders as they move in this direction.

I. CHANGING TIMES REQUIRE MORE POWERFUL RESPONSES

Given our nation’s proud history of pragmatism and ingenuity, our efforts to become more effective in addressing today’s most serious social problems shouldn’t be stuck with outmoded problem-solving structures. In the face of multiple obstacles that stand in the way of the progress we all want to see, we suggest that **we can overcome one significant obstacle by expanding our understanding of what constitutes useful and usable evidence.**

In this paper we explore how we, as a nation, can move forward by becoming smarter in learning from our extensive knowledge and experience in order to improve results and to assure the responsible expenditure of public and philanthropic funds.

The question we address is not whether we need evidence, or even whether we need more evidence. The question we explore here is: **How do we obtain and apply the kinds of evidence that will be most useful in achieving significantly greater outcomes?**

We need a new evidence framework to respond to changing times: we are tackling tougher problems, the results of many past interventions have been disappointing, and our new ways of working require new and continuous ways of learning and generating knowledge that can, in turn, lay the foundation for further progress. We identify several efforts now underway to respond to changing circumstances (Section II). We then describe the components of a more robust evidence framework that will produce the full spectrum of credible evidence that counts when it comes to improving results (Section III).

In Section IV of the paper, we identify five elements of a powerful approach to evidence. We describe the work of leaders at many levels and in a variety of domains who are operationalizing these elements in several diverse initiatives. Increasingly, these initiatives base their work on a variety of sources of knowledge; creatively deploy results-based approaches, continuous-improvement methods, and networked learning communities; use varied evaluation methodologies to assess what they do; and put in place an infrastructure for learning that allows them to adapt and improve interventions over time and to achieve results at greater scale.

The individual elements of the more powerful approach to evidence that their work illustrates are not new; all have been identified and examined in existing bodies of knowledge, and each is currently in use somewhere. **What we call for, and what has not yet happened, is the intentional and systematic creation of the conditions under which these diverse ways of generating knowledge can be coherently implemented, rigorously examined, and supported over time.**

We conclude by raising the question of what it would take for the pragmatic, results-based approach to evidence described in this paper to become routine, the “new normal.” The 2014 Harold Richman Symposium is a forum to begin to answer that question.

WE ARE TACKLING TOUGHER PROBLEMS

The problems we are addressing have become increasingly complex and more entrenched. We seek to reduce disparities in health, education, and other population outcomes that are based on race, income and class. We seek to keep our nation from becoming less competitive economically and less cohesive socially. We seek better lives for our children even as the nation is experiencing more income inequality, less upward mobility, larger numbers of single-parent families and greater numbers of people living in census tracts where at least 40% of residents are at or below the federal poverty line. The characteristics of these neighborhoods, including not only poverty but also crime, violence, dilapidated housing, high unemployment, poor schools and few social supports, are all risk factors for poor long-term outcomes. Perhaps most shocking, and galvanizing to action, is that the population of these high-risk neighborhoods increased by 72% between 2000 and 2012.¹

Because of the multiply determined nature of the problems we face, and because we have lagged in developing more powerful interventions, we should not be surprised to find that the results of many of our past efforts have been disappointing.

EARLIER RESULTS HAVE BEEN LIMITED

The results of too many interventions have been disappointing in that they (a) didn't respond to the children and families with the greatest needs, (b) had only marginal impact on those they did reach, and/or (c) failed to reflect what we are learning from research and experience about the best ways to intervene, either directly or indirectly through policy and environmental changes.

Many programs do not help the highest needs populations. A review of early childhood interventions by Jack P. Shonkoff, director of Harvard University's Center on the Developing Child, found that large numbers of young children and families who are at greatest risk of poor outcomes do not appear to benefit significantly from existing programs—particularly those children who experience toxic stress associated with persistent poverty complicated by racial discrimination, child maltreatment, maternal depression, parental substance abuse, or interpersonal violence.²

Similar evidence comes from the Social Genome Model,^A which allowed Isabel V. Sawhill, co-director of the Brookings Institution's Center on Children and Families, and colleagues to simulate long-term effects of childhood interventions. These scholars explored what would happen if all low-income children and youth were to participate, during early and middle childhood, in a succession of proven programs that have strong empirical track records of improving outcomes for lower-income participants.^B Sawhill and colleagues found that large disparities in white-black well-being still persist even after the idealized interventions.³ The greatest progress occurred among individuals in the middle economic group, not the most disadvantaged. For the latter group, the post-intervention white-black success gap remains at 14% in middle childhood, 22% in adolescence, and 29% in early adulthood.⁴

^A The Social Genome Model was originally developed by Isabel Sawhill at the Brookings Institution and is now based at the Urban Institute as a collaborative effort of the Brookings Institution, Child Trends, and the Urban Institute.

^B The programs used in the simulations were PALS (Play and Learning Strategies), "quality preschool," Success for All, a "strong Social Emotional Learning program" and small high schools of choice.

Many programs are found to have only marginal impact on those they do reach. For example:

- Among 90 randomized-control trials (RCTs) commissioned by the Institute of Education Sciences since 2002 to evaluate the effectiveness of diverse educational programs, practices, and strategies, 90 percent found weak or no positive effect.⁵ Similarly, of the 13 RCT-evaluated interventions by the U.S. Department of Labor that reported results since 1992, about 75 percent were found to have had weak or no positive effects.⁶
- The Shonkoff review cited above found that the most-effective interventions for children living in poverty can achieve statistically significant impacts on long-term outcomes, but the size of their effects often are modest, and persistent disparities remain.⁷

There are multiple possibilities as to why so many interventions seem to have modest or no effects. Among the reasons is that past evaluations have often failed to distinguish between high-quality and poor implementation. In addition, the interventions examined by the studies cited above focused predominantly on stand-alone programs targeted on individuals that lend themselves to assessment by experimental methods and are most readily scaled. It is also possible that the design of many of these programs had failed to take into account the contextual variability of the settings and circumstances in which they would be implemented.

Not enough of what we know through research and experience is reflected in today’s practices and policies. For example, we possess but do not always apply knowledge about the importance of adult-child relationships to early child development. Many policies still define “quality” in early child care and education in terms of adult-child ratios, group size, and physical facilities rather than ensuring “that relationships in child care are nurturing, stimulating, and reliable, [leading to] an emphasis on the skills and personal attributes of the caregivers, and on improving the wages and benefits that affect staff turnover.”⁸

To take another example, the Center for Juvenile Justice Reform at Georgetown University, in a review of 548 programs aimed at reducing recidivism among delinquent youth, found that work with juvenile offenders was much more likely to be effective and to reduce recidivism when the intervention took a therapeutic approach to changing behavior rather than a control or deterrence philosophy⁹—a finding still often not reflected in practice and policy.

PREVAILING WAYS OF OBTAINING EVIDENCE ARE MISMATCHED TO NEW WAYS OF WORKING

Randomized-control trials (RCTs) are and have been the appropriate and best evaluation method to assess circumscribed interventions that could be held constant over time and over diverse sites, and that targeted well-defined goals for problems that had well-defined and direct causes. The ability of RCTs to provide clear information about what *can* work and what hasn’t worked, at least within their experimentally controlled settings and circumstances, has been responsible for considerable progress. Their findings build support among policymakers, funders and the public for important programmatic investments. The Nurse Family Partnership (NFP) is an example: The three randomized-control trials conducted by NFP on three different populations in the 1980s and 1990s, provided a major impetus to putting home visiting on the national social policy agenda. Without them—ultimately supplemented by evidence from other strong national models—we might not be seeing a \$1.5 billion federal investment

in home visiting today. We also would not be seeing some of the enthusiasm for public investment in support of young families by policymakers and the media, which is so often based on the clear findings of effectiveness of the NFP intervention.^C

In the current era of scarce resources, diminished trust in government and large institutions, and the “the fog of confusion” about what is worth investing in,¹⁰ widespread reliance on the clarity of RCTs and other experimental evidence seems to make good sense. Knowing what works from experimental evidence removes considerable risk from decisions about resource allocation, accountability, and the selection of models for implementation, funding, and scaling interventions. And it protects decision makers at all levels from accusations of bias.

The logic of an assessment method that removes doubt about the link between an intervention and its results was reinforced as we became accustomed to relying on the results of testing drugs and other medical interventions this way. By the late 20th century, RCTs were seen as the optimal method for “rational therapeutics” in medicine. Although the analogy between discovering what works in medicine and in social policy was imperfect,^D the confidence in RCTs and other experimental evidence extended to the social sphere, and randomized experiments were widely recognized as the “gold standard” for evaluation.

But there were unintended consequences in relying on RCTs as the best way of assessing every kind of intervention. The flaws in an evaluation hierarchy with RCTs at the undisputed top, regardless of the questions that needed answers, were becoming apparent. Some critics thought RCTs were too expensive and produced results too slowly. Respected researchers pointed out that the controlled conditions in which they were conducted could not reflect the wide range of conditions in which the interventions would be scaled, and provided little guidance on real-world, real-time implementation. Additional concerns came from the communities, agencies, practitioners, and reformers that were engaged in new ways of addressing old problems. The change strategies they employed were often complex and interactive, characterized by attributes in Figure 1.

In response to changing problems, limited results, and new ways of working we have seen two sets of recent developments. The first, described in Section II, calls for a greater role for evidence and new tools to achieve that goal. The second, which we describe in Section III, we see as building a sturdier evidence framework.

^C Most recently, Nicholas Kristof and Sheryl Wudunn wrote in the *New York Times Sunday Review* (9-12-14) under the headline, “The Way to Beat Poverty,” that NFP was “stunningly effective.” They call on donors to support this “anti-poverty program that is cheap, is backed by rigorous evidence, and pays for itself several times over in reduced costs later on.”

^D Randomized trials in medicine tended to involve interventions that were well-defined, standardized, and circumscribed, featuring a biological organism that was relative homogeneous across individuals and populations. This did not prevent many economists and policy makers from seeking to apply the same methods to identifying “what works” in domains for which interventions were less well-defined, less likely to be circumscribed, and where standardization often undermined effectiveness in the heterogeneity of the target populations in which they were implemented.

**Fig. 1:
COMMON ATTRIBUTES OF COMPLEX INTERVENTIONS**

- They are not targeted only on individuals, but aim to change families, neighborhoods, larger communities, norms, and systems.
- They pursue multiple, intertwined, and even unanticipated goals.
- They involve collaboration across professional and organizational boundaries.
- They seek to reform systems to respond to the complex environments and cultures that influence outcomes.
- They rely on subtle and hard-to-measure effectiveness factors, including trusting and respectful relationships.
- They deal with issues not inherent within a program model.
- They devise solutions uniquely suited to their particular time, place, and participants.
- They integrate proven and promising practices with ongoing activities.

II. EARLY RESPONSES

Three recent developments in the evidence arena are variations that improve on prevailing practices. They: (1) call for greater use of evidence in making major decisions; (2) encourage the development and support of faster, cheaper randomized trials; and (3) use tiered hierarchies of evidence to make funding decisions.

CALLS FOR GREATER USE OF EVIDENCE

Among the most widespread concerns about the role of evidence in decision-making has been the weak role that any kind of evidence has played, especially when compared to the influence of ideology, politics, history, and even anecdotes. University of Chicago child welfare researcher Deborah Daro observes that merely getting research used at all can seem like a victory, given how evidence of any kind plays such a small part in most decisions in both the public and philanthropic sector. But there are signs, reports Robert Shea of the [National Academy of Public Administration](#), that governments at all levels are putting more emphasis on the need for evidence, on linking funding to evidence-based practices and reforms, on supporting performance management, and on persuading members of Congress to “embrace the evidence agenda.”

Much of the recent increased interest in the greater use of evidence in governmental decisions at the federal level has been propelled by [Results for America](#). In October, 2013, it opened a campaign called [Moneyball for Government](#), signing up public officials and opinion leaders to encourage government officials and others to base future investments on [three principles](#):

- Building evidence about the practices, policies and programs that will achieve the most effective and efficient results so that policymakers can make better decisions;
- Investing limited taxpayer dollars in practices, policies and programs that use data, evidence and evaluation to demonstrate they work; and
- Directing funds away from practices, policies, and programs that consistently fail to achieve measurable outcomes.

The campaign's materials open with a statement that "less than \$1 out of every \$100 of government spending is backed by even the most basic evidence that the money is being spent wisely." In an article in *The Atlantic*, Moneyball signatories and former government officials Peter Orszag and John Bridgeland wrote they "were flabbergasted by how blindly the federal government spends," and by how other types of American enterprise make sophisticated spending decisions while the federal government's "spending decisions are largely based on good intentions, inertia, hunches, partisan politics, and personal relationships."¹¹ The challenge now is not only to increase dramatically the use of evidence in governmental decision-making, but also to expand the available evidence to better match the requirements of more complex interventions in systems and communities.

DEVELOPMENT OF FASTER, CHEAPER RANDOMIZED TRIALS

Experiments with low-cost, large-scale randomized control trials are being advanced by the White House Office of Science and Technology Policy in partnership with the Coalition for Evidence-Based Policy.¹² The premise is that because RCTs, when well designed and carefully implemented, can assure that any difference in outcomes between an intervention's treatment and control groups can be confidently attributed to the intervention and not to other factors, sizable RCTs that can be done relatively quickly and at low cost could be "an important complement to traditional, more comprehensive RCTs as part of a larger research agenda."¹³ The strategy for fielding low-cost, sizable RCTs is based on (a) embedding random assignment into initiatives that are being implemented anyway as part of usual program operations, and (b) using administrative data that are collected already for other purposes to measure the key outcomes.¹⁴

Low-cost rapid RCTs fit well with recently developed financing options such as: Social Impact Bonds (SIBs)^E and the idea of "paying for success." Both aim, according to the Federal Reserve Bank of San Francisco's review, "to transform the social sector into a competitive marketplace that efficiently produces poverty reduction by holding the social sector accountable through ex-post payments for evidence-based results rather than ex-ante payments for promising programs."¹⁵

^E With Social Impact Bonds, the government sets a specific, measurable outcome that it wants to achieve in a population and promises to pay an external organization for accomplishing the outcome. Private investors provide the working capital to hire and manage service providers and a third-party evaluator. If the stipulated outcome is achieved, the government releases an agreed-upon sum to the external organization, which then repays its investors with a return for assuming the upfront risk. If the agreement fails, the government is not on the hook to pay, and investors are not repaid with public funds. See <http://www.americanprogress.org/issues/economy/report/2014/02/12/84003/fact-sheet-social-impact-bonds-in-the-united-states/>.

Both depend on undeniable evidence of short-term success or failure of the circumscribed intervention, which makes rapid, low cost randomized trials a useful tool.

APPLYING TIERED HIERARCHIES OF EVIDENCE TO FUNDING DECISIONS

When federal agencies have used evidence to make funding decisions in the last decade, they have tended to apply a tiered hierarchy, at the top of which is evidence from randomized-control and quasi-experimental studies.

Major federal initiatives that currently prioritize funding for complex interventions using experimental evaluations include:¹⁶

- [Investing in Innovation Fund \(i3\)](#) at the Department of Education;
- [Teen Pregnancy Prevention Initiative](#) and the [Maternal, Infant and Early Childhood Home Visiting Program \(MIECHV\)](#) at the Department of Health and Human Services;
- [Social Innovation Fund \(SIF\)](#) at the Corporation for National and Community Service;
- [Trade Adjustment Assistance Community College and Career Training \(TAACCCT\)](#) at the Departments of Labor; and
- [Workforce Innovation Fund](#) at the Departments of Labor, Health and Human Services and Education.

Across departments, the definitions of what counts as “evidence-based” are similar but not identical: The highest ratings and the most funds go to initiatives that have been found effective by studies using a randomized-control trial or quasi-experimental design. Lesser ratings may go to randomized-control trials that are good but fall short of “strong evidence” and to comparison-group studies in which the intervention and nonrandomized comparison groups are very closely matched.

The use of a tiered evidence framework reflects the reigning scientific paradigm of the past few decades, which asks a simple question: Does independent variable A (the intervention) cause the desired change in dependent variable B (the outcome)? A primary reliance on this one method of obtaining knowledge privileges single-level programmatic interventions. These are most likely to pass the “what works?” test within the controlled conditions of the experimental evaluation. But the programs so identified may not be the most effective and scalable interventions to improve lives, and they often do not produce the same results under real-world conditions. Reliance on this hierarchy also risks neglecting or discouraging interventions that cannot be understood through this methodology and sidelining complex, multi-level systemic solutions that may be very effective but require evidence-gathering methods that rank lower in the evidence hierarchy. Consequently, our progress in solving major social problems has often been undermined by pressure to adhere to the constraints of this hierarchy.

These steps toward more, and more usable, evidence represent in different ways a broadening of approaches to define evidence. With Results for America and others, we believe that much more needs to be done, however, to secure the actionable, real-world knowledge and the climate of learning now needed to achieve substantially better outcomes, more effective use of resources, and greater accountability.

A growing number of key observers and decision-makers are pointing to the importance of this shift in how we think about evidence. For example:

- The federal Office of Management and Budget (OMB) no longer insists that only RCTs can produce credible evidence to guide policy and practice. On July 26, 2013, the OMB, Domestic Policy Council, Office of Science and Technology Policy, and Council of Economic Advisers jointly issued a memorandum calling on the heads of federal departments and agencies “to continually improve program performance by applying existing evidence about what works, generating new knowledge, and using experimentation and innovation to test new approaches to program delivery.” The memo also encourages agencies to develop “strategies to use evidence to promote continuous, incremental improvement” and to submit proposals “that would test higher-risk, higher-return innovations with the potential to lead to more dramatic improvements in results or reductions in cost.”¹⁷
- Philanthropists are incorporating new perspectives based on insights about collective impact, awareness of the connections that span disciplines and jurisdictions, a disillusionment with linear models, and a recognition of the importance of context and complexity. Describing evolving perceptions within philanthropy, John Kania, Mark Kramer, and Patti Russell conclude that funders cannot afford “to overlook the complex dynamics and interpersonal relationships among numerous nonprofit, for-profit, and government actors that determine real world events.” They advocate a focus “on learning and sensing opportunities, not just on evaluating the outcomes attributable to specific interventions,” because outcomes do not “arise from a linear chain of causation that can be predicted, attributed, and repeated.”¹⁸

III. BUILDING A STURDIER EVIDENCE FRAMEWORK

Encouraged by a new openness to deeper, more practical and more flexible ways of determining what constitutes progress in improving outcomes and accountability, we now describe a second set of recent developments: the sturdier evidence framework being created around the country. The components of this evidence framework are summarized in Figure 2.

We see this evidence framework as giving us *more and better* information—if not always *more-certain* information.

This is the evidence that will lead to the progress we are looking for and that will shape an expanded understanding of what *evidence-based* means. Qualifying as *evidence-based* does not depend on an intervention’s having been validated by a single method. To be solidly evidence-based, decisions about programs, policies, strategies, and practices must be the product of multiple efforts to understand them. Expanding the definition of *evidence-based* allows us to fit the tool to the task—using a hammer to drive in a nail, and a wrench to turn an object. The sturdy evidence framework we now describe assumes an expanded set of tools that make possible this expanded definition of what it means to be *evidence-based*.

**Fig. 2
COMPONENTS OF A
STURDIER EVIDENCE FRAMEWORK**

1. Systematic learning is the foundation
2. Accountability for results is an essential pillar
3. Useful and usable measure are made more readily available
4. Complexity is fully recognized
5. A full range of evidence is mobilized in decision making

1. SYSTEMATIC LEARNING IS THE FOUNDATION

Systematic learning, going well beyond thumbs-up or thumbs-down determinations of what has worked in the past, illuminates what is likely to work in the future—including how it works, why, for whom, in what circumstances, and at what scale. This deeper, broader, integrated approach to learning helps us understand why a particular result occurs, and in doing so promises to produce the “relevance” and “usefulness” that John Q. Easton called for when he became director of the Department of Education’s Institute of Education Sciences in 2009.¹⁹

A systematic learning stance helps stakeholders implement evidence-based programs (EBPs) when those are appropriate, but also prevents an over-reliance on EBPs. Relying on EBPs even when they are a poor fit usually happens in the hope of minimizing the risk of funding or deploying ineffective interventions.^F A systematic learning stance helps in the identification of effectiveness factors, which make it possible to spread programs to settings where success will depend on adaptation.

Effectiveness factors, also called **core components**, are shared by successful interventions that pursue similar goals. Implementation experts Karen Blase and Dean Fixen describe these core components as “the functions or principles and related activities necessary to achieve outcomes”²⁰; they may be programmatic, or they may involve implementation or contextual conditions. The process of determining precisely which aspects of a successful intervention are effectiveness factors requires good data and thoughtful judgment.

Effectiveness factors are portable. They are often the best guide to action, and have a multiplier effect, because they are more transferable to other settings than model programs and because their identification makes thoughtful adaptations possible. When effectiveness factors have been identified, it becomes possible for the people working to design, improve, and spread successful interventions to implement EBPs when those are appropriate, but also to adapt them and develop new interventions that incorporate the effectiveness factors. As noted later in the paper, complex interventions often use EBPs as a starting point but then develop, implement, and test the additional strategies needed to achieve the outcomes they are seeking to achieve.

Systematic learning, with the deeper understanding of the effectiveness factors it leads to, will allow—indeed encourage—adaptations of proven models. Information about effectiveness factors can be applied to new and existing programs to strengthen their design. Implementers would not be required to hold every aspect of an evidence-based program constant, applying the experimentally tested intervention with “fidelity”—a strategy with limited success. Rather, funders can encourage people on the ground to continually improve outcomes by adapting to the needs of particular localities, cultures, and populations; by rapidly incorporating advances in knowledge, changes in context, new data, lessons from practice, and findings from evaluations; and by testing these so they might learn as they go. This line of thinking suggests that we consider “fidelity” in new ways—not by insisting that grantees implement *models* with fidelity but by looking for adherence to core principles, making lessons

^F This hope is based on three flawed assumptions: that (1) “proven” programs are permanent solutions to problems that are assumed to be static; (2) any evidence-based program is a good fit with the needs and strengths of any other specific community or system; and (3) the variables held constant in an intervention and control groups and the selection of consenting participants and their varied dropout rates won’t misrepresent or undo the effectiveness of the intervention when implemented beyond the research setting.

from implementation in varied contexts available to the wider field, and encouraging learning that leads to adaptations that maximize impact.⁶

We see some movement to support thoughtful adaptation in a 2013 research brief from the U.S. Department of Health and Human Service’s Assistant Secretary for Planning and Evaluation, which suggested, “It is possible to adapt an evidence-based program to fit local circumstances and needs as long as the program’s core components, established by theory or preferably through empirical research, are retained and not modified.”²¹

When effectiveness factors are identified across interventions, efforts to scale impact are strengthened, according to Child Trends President Carol Emig. She has found that understanding the elements of effectiveness is a good way to improve programs:

Instead of telling a city or foundation official that they have to defund their current grantees because they are not evidence-based, funders can tell long-standing grantees that future funding will be tied at least in part to retooling existing programs and services so that they have more of the elements of successful programs.²²

For the same reason, Grantmakers for Effective Organizations (GEO) encourages foundation leaders to gain insights into the commonalities of various solutions, not only among their own grantees but also among other funders.²³

Systematic learning takes into account the context within which interventions can thrive. Contextual success factors are particularly important to scale-up efforts, since so many factors necessary for success are not inherent in program models. These contextual success factors include:

- The neighborhood and community context, including its full range of simultaneous and interacting challenges, demands, stresses, priorities, and resources;
- The composition of the target population;
- The capacity of the organization implementing the program and its relationship to the target population;
- The availability of a well-trained, motivated workforce; and
- The availability of funding under conditions that support and don’t undermine the intervention’s operations.

⁶ The Nurse Family Partnership, cited earlier as an example of the benefit of clear findings of effectiveness from randomized trials, has also become an example of a program that may be too rigid in insisting on fidelity in implementation, even though the model may not be a good fit for every community. As Haskins, Paxson, and Brooks-Gunn point out, programs that work for first-time teen mothers may not be suitable for parents with drug addiction or serious mental health issues. (See Haskins, R., Paxson, C., and Brooks-Gunn, J. Fall 2009. “Social science rising: A tale of evidence shaping public policy.” *Future of Children Policy Brief*, http://futureofchildren.org/futureofchildren/publications/docs/19_02_PolicyBrief.pdf.) Yet local agencies are often discouraged from making adaptations when they are unable to find enough nurses to make the home visits or when they seek to expand the original model to serve clients who had received no prenatal care, suffered from severe depression, were involved with substance abuse or domestic violence, were expecting a second (or third or fourth) child, or were threatened with homelessness. In response to such concerns, NFP founder, David Olds, and colleagues currently conduct research focused on quality improvement, model improvement, and implementation in NFP practice settings. (See http://pediatrics.aappublications.org/content/132/Supplement_2/S110.full.html)

Patrick McCarthy, president of the Annie E. Casey Foundation, noted the importance of context when he declared that “an inhospitable system will trump a good program—every time, all the time.”²⁴ Srik Gopalakrishnan of FSG concurs, writing that because “context can make or break an intervention,” it is not tenable to assume that a program once shown to be effective through an RCT will always remain effective, irrespective of context.²⁵

“Systematic learning for improvement,” developed originally by the Institute for Healthcare Improvement and now gaining traction around the country, is at the heart of the work promoted by Tony Bryk of the Carnegie Foundation for the Advancement of Teaching. Bryk calls it “improvement research”—a practice that culls and synthesizes “the best of what we know from scholarship and practice,” rapidly develops and tests prospective improvements, deploys learning about what works, and adds to the knowledge that can continuously improve the performance of the system.

Using networks established for this purpose, the Carnegie Foundation is orchestrating “a common knowledge development and management system to guide network activity and make certain that whatever we build and learn becomes a resource to others as these efforts grow to scale, [offering] a prototype of a new infrastructure for research and development.”²⁶

Learning for improvement emphasizes multiple, rapid tests of small changes tried by varied individuals working under different conditions.¹ Each test provides a bit of evidence and a bit of local learning. Bryk has found that by organizing tests of small changes around causal thinking, linking hypothesized solutions to rigorous problem analysis and common data, it is possible to accelerate learning for improvement at scale.²⁷

Taking this learning approach means “being open to testing not only your precise solutions but your understanding of the problem you are addressing. As you experiment, you [may be forced]...to question whether you have articulated the right goal or misjudged what it will take to succeed,”²⁸ observes Katherine Fulton, president of the Monitor Institute. One must be willing to confront “messy political, social, and human realities, with a strong tolerance for ambiguity and a keen eye for the dynamic contextual forces that can be harnessed for your ends,” she concludes.²⁹

Lastly, observers have speculated that a culture of continuous learning for improvement may actually do more than improve results, but may maintain, over time, the passion that people started with when they entered the helping professions.

²⁴Lisbeth Schorr came to a similar conclusion in seeking to understand why so many of the successful programs described in her 1988 book *Within Our Reach* had disappeared, or had been diluted, dismembered, or destroyed just five years later. Once they came out of the hothouse and were no longer under the protective bubble of a “demonstration program,” the real-world contexts in which they had to operate undermined them. See *Common Purpose* (1997).

¹ “Small controlled experiments, which may include randomization, are a healthy part of any innovation and development process,” especially when it comes to a single, simple variable, Srik Gopalakrishnan has observed. “For example, through simple testing, the 2012 Obama campaign discovered that the most successful email subject line (in terms of getting campaign contributions) was simply, “hey.”

<http://www.fsg.org/KnowledgeExchange/Blogs/StrategicEvaluation/PostID/478.aspx>

2. ACCOUNTABILITY FOR RESULTS IS AN ESSENTIAL PILLAR

No one today would argue that efforts to improve outcomes, no matter how nobly conceived, can neglect demands for evidence of effectiveness. The evidence of good stewardship of investment must go to funders and all other stakeholders, including individuals and families most directly affected by the interventions.

Demonstrating accountability for results is not in conflict with learning for improvement. Each supports the other. Systematic learning gets data to actors faster and creates a culture in which they accept responsibility for error and failure and for making changes necessary for improvement. It is part of a shift away from a compliance-driven culture to a learning organization, and from “gotcha evaluations” to continuous learning and improvement.

Networked communities seem to be on the way to achieving a real-life integration of accountability and learning for improvement. Networks “design and test effective interventions while generating learning about how these work, for whom, and under what organizational conditions,” explains Anthony Bryk.³⁰ In a diverse network of sites, specific, measurable outcomes create shared targets; a coherent and explicit chain of reasoning guides intervention design; and, in each locale, problem solving occurs simultaneously with inquiry into whether the changes are improvements.³¹

Even though the presumed conflict between accountability and learning has been resolved more often than not, there are still instances in which they seem to involve irreconcilable choices. The authors of this paper hear frequently from program managers and staff about their frustration when they are asked to resist making changes based on lessons learned from their daily experiences. Sometimes they are told that adaptations would threaten the rigor of the evaluation. Others find themselves held accountable not for the outcomes they achieve, but for the fidelity to the original design with which they are implementing their program even when ongoing learning during implementation makes it clear that outcomes will not be achieved unless they make an adaptation in the intervention model. The assumption has been that “implementer discretion can undermine the pursuit of policy objectives.”³²

It would be a mistake to allow a presumed conflict between accountability and learning to stand. It is worth struggling to reconcile these perspectives. There is, in fact, increasing recognition of the interdependence between accountability and learning for improvement. FSG’s *Guide to Evaluating Collective Impact* calls for a “shift to using data in the service of learning *and* accountability.”³³ Similarly, Irene Guijt, a leader in international participatory learning, writes:

*You cannot be accountable if you do not learn. And you need to know how well you live up to performance expectations in order to learn. The tug-of-war between learning and accountability is nonsensical. They need each other. Understanding effectiveness requires both.*³⁴

It is easier to call for reconciling accountability with continuous learning than to do so, however. Tony Bryk urges adopting and maintaining an improvement orientation, which asks “Where are we succeeding, where are we not and how can we get better at the latter?” This view of accountability is consistent with that called for by political scientist Arthur Lupia, who suggested in his 2013 plenary address to the American Evaluation Association that “accountability needs to shift from achieving predetermined results on a predetermined plan to demonstrating the capacity to achieve results in dynamic environments.”³⁵

As the field gets better at defining the interim markers that research has shown will lead to desired longer-term results, some of the burden of each intervention having to demonstrate the connection anew will be lightened.

3. USEFUL AND USABLE MEASURES ARE MADE MORE READILY AVAILABLE

For many of the most ambitious recent efforts to improve outcomes, the appropriate measures to assess progress are not readily available—or simply don't exist. **New measurement tools are needed that more accurately account for critical variations in locale, ways of working, and objectives, and that actively engage providers and clients in using them and applying the findings that emerge.** For communities and reformers increasingly turning to more collaboration and more complex interventions, the difficulties are exacerbated as they try to find measures of success that fit their new ways of working. Agreeing on shared goals and defining success is hard enough when partners with different worldviews and mandates begin to work together, but finding *measures of success* they can agree on is even harder.

Because available measures so often don't match measurement needs, the availability of specific but not necessarily useful or appropriate measures frequently drives the choice or design of interventions instead of the other way around.

When Anne Kubisch headed the Aspen Roundtable on Community Change, she observed that people aiming for change at the population, institutional, community, and systems (policy) levels often have good theories but no corresponding measures. In particular, they lack the tools to measure community capacity, cohesion, civic capacity, and systems and policy change—let alone tools that would convert such measures into evidence of cost-effectiveness. As a result, Kubisch told us:

Interventions that might aim for change at the population, institutional, community, and systems level, get dumbed down or over-simplified, trading off the opportunity to achieve breakthrough outcomes for the prospects of being able to document effectiveness.³⁶

The measurement challenge posed by place-based efforts is matched by the challenge at the other end of the spectrum, by the initiatives that carefully tailor their work to the specific situation and trajectory of each participant. Their measurement activities are complicated because they allow program activity to evolve over time and within a community context,³⁷ and because they take an asset-based approach to communities and clients. Katya Fels Smyth, founder of the Full Frame Initiative (FFI), which works with public agencies and nonprofits to better help people at the “deep end of the deep end”, explains that they don't see clients and their families and communities as initially “broken,” only to be “fixed” by participating in a program. She writes that work with those living at the intersection of poverty, violence and trauma is hampered by a lack of tools for assessing effects on participants' safety, stability, and social connections. Smyth and colleagues contend that prevailing measures of success target only problems that have specific, well-defined solutions and “devalue the long-term and often rich and nuanced impact of full-frame, deep, flexible, comprehensive approaches,” that encourage a

client to become part of “a community in which the individual is rooted in something bigger and broader than herself and her problems.”^{38, J}

Some of these measurement challenges require the development of new measures, both locally and nationally. Some will be best dealt with when local organizations, with their fine-grained understanding of specific measurement needs, collaborate with outside measurement experts to develop measures that reflect and address varying concerns and characteristics that vary by locality and context. And we need increased capacity locally for developing and using these.

Some measurement challenges can be responded to nationally. Among national organizations trying to address the lack of appropriate measures, Child Trends is probably the leader. It has identified (a) measures of positive well-being among infants, toddlers, and older children aged 6 to 17; and (b) measures of the quality of early care and education settings for policy-related purposes. In addition, the Child Trends DataBank promotes “the efficient sharing of knowledge, ideas, and resources within the child and youth indicators community” and examines and monitors more than 100 indicators of risks and positive development for children and youth.³⁹ In addition, the Urban Institute is about to issue a new book, Strengthening Communities with Neighborhood Data, documenting how government and nonprofit institutions have used information about neighborhood conditions to change the way we think about community and local governance in America.

Outside experts in universities, evaluation and advocacy organizations, and even businesses will be needed to help local practitioners acquire capacity to use emerging collections of data (including new electronic databases) in innovative and economical ways for planning and analysis to enhance individual, family, and community well-being.

4. COMPLEXITY IS FULLY RECOGNIZED

We see a growing recognition that addressing complex problems requires complex responses. Reformers and practitioners are acknowledging the growing complexity of their efforts in response to the complexity of problems they are addressing. Some find that complexity gets in their way as they try to connect activities to results and establish clear lines of responsibility for outcomes among multiple partners in several domains. A tempting response is to define complexity away. But **if complexity is not explicitly taken into account by obtaining and using multiple streams of evidence, the risk is that we distort investments, stifle innovation, and reduce the opportunity to achieve impacts at scale.**

Complex problems are characterized by Kania, Kramer, and Russell as dynamic, nonlinear, and counter-intuitive, the result of the interplay between multiple independent factors that create a kaleidoscope of causes and effects that can shift the momentum of a system in one direction or another in

^J Contrasting their approach to those that are more prevalent, they write that in the latter, *A client or patient’s own complex and messy understanding of her situation is often subjugated to a provider’s characterization. The program may push a woman to understand her situation in terms consistent with those identified by the funder so that the program can demonstrate “success” at producing specific outcomes, irrespective of whether those outcomes fit within the woman’s understanding of her larger situation, or leave her with losses in other arenas.* (See: Smyth, K.F., Goodman, L., and Glenn, C. October 2006. “The Full-Frame Approach: A New Response to Marginalized Women Left Behind by Specialized Services.” American Journal of Orthopsychiatry 76(4):489-502.)

^K “Stacking EBPs” is the formulation of our colleague, Marian Urquilla.

unpredictable ways.⁴⁰ They use the example of population health to point to the factors increasingly thought of as *social determinants*, which include the availability and quality of health care, economic conditions, racial discrimination, social norms, daily diet, inherited traits, familial relationships, weather patterns, and psychological well-being.

Katya Fels Smyth of the Full Frame Initiative points out an additional dimension to consider: that complex problems typically exist in *conditions* of complexity. She and her colleagues write about “the wildly unpredictable environments in which people live...as well as the social history of race, class, gender, and place baked into community, personal identity, behavior, and worldview.”⁴¹ On the front lines today (and earlier, in academia), these complex situations have been called “wicked problems”—usually ill-defined and often morphing over time and in response to changing contexts.⁴²

It is hard to acknowledge and address complexity. Doing so requires more than simply adopting an ever-expanding number of variables or arrays of statistical tests. Rather it requires investing in capacity, infrastructure and resources that permit ongoing, data-driven change and learning over time, and that draw in all stakeholders to work together and share ownership and accountability. This approach does not jettison programmatic interventions. But when communities and funders choose the interventions they will implement or support, they do not rely primarily on lists of EBPs or “proven programs,” knowing that “stacking EBPs”^k does not result in a coherent and responsive set of interventions. Rather, they carefully select those that respond to the actual problems that their initiative is addressing, adapt these to local contexts, and link them to each other and to systemic interventions. They supplement these by building on what they can learn about effectiveness factors to design additional ways to achieve agreed-upon outcomes.

Recognizing complexity and its implications also means:

- Developing and using an array of innovative ways to analyze and generate evidence about what’s worth implementing, supporting, and spreading;
- Resisting the yearning to simply replicate prior interventions, and struggling instead to build on and adapt what has been shown to work, including how different effectiveness factors come together and become meaningful in particular contexts; and
- Acknowledging that many promising interventions are sufficiently complex that they cannot be studied as controlled experiments and are not appropriate for experimental or quasi-experimental research designs (although some components of them may be)^l. The notion of “natural experiments” applies to complex interventions only loosely.

^k “Stacking EBPs” is the formulation of our colleague, Marian Urquilla.

^l Simon Cohn of the University of Cambridge’s Behavior and Health Research Unit writes in the *Journal of Health Services Research Policy*, “A commitment to the generalizability and reproducibility of things that are inherently complex is naïve and misplaced. A richer appreciation of complexity and the commitment to the RCT as the gold standard of evidence are ultimately incompatible.”

^l Simon Cohn of the University of Cambridge’s Behavior and Health Research Unit writes in the *Journal of Health Services Research Policy*, “A commitment to the generalizability and reproducibility of things that are inherently complex is naïve and misplaced. A richer appreciation of complexity and the commitment to the RCT as the gold standard of evidence are ultimately incompatible.”

Along similar lines, Thomas Kelly, Jr., vice president for knowledge, evaluation, and learning at the Hawaii Community Foundation, warns that when it comes to community change initiatives “too many contextual variables are present for evaluators to control for or [to] examine fully...the rich fabric of community life, and all the

Funders often are tempted to shrink from the implications of recognizing complexity. Two groups have studied the ways in which funders deal with complexity: Patricia Patrizi, Elizabeth Heid Thompson, Julia Coffman, and Tanya Beer of the Evaluation Roundtable⁴³, and FSG’s Kania, Kramer, and Russell.⁴⁴ Both groups focused on philanthropy, but many of their findings and insights extend to public funders as well. For instance, Patrizi et al. reported that foundation strategy is hampered by a failure to recognize and engage with complexity and that many funders prefer to frame their strategies as theories of change with a set of linear, causal, and certain actions. In the face of the “often mind-boggling complexity of this work,” many go to great lengths to avoid attempting to understand or account for the complexity in their work and to distance themselves from the “mess” of social change.⁴⁵

Darren Walker, president of the Ford Foundation, may be a positive outlier in this respect. He has argued that we can “embrace complexity without neglecting rigor and outcomes.” He calls on us to “open our eyes and minds to the entire spectrum of alternatives—approaches that anticipate and embrace complexity” in order not to miss potential breakthroughs as “we look for silver bullets and simple solutions.”⁴⁶

To embrace complexity without neglecting rigor and outcomes requires a learning mindset. In the words of evaluator Michael Quinn Patton, “Rigor does not reside in methods, it resides in thinking.” Patrizi and her colleagues suggest that taking complexity seriously means making a commitment to “learning from action,” a variation of learning for improvement. Here, too, learning builds on cycles of acting, making sense of what happened, and drawing implications for further action. Patrizi et al. state that deep understanding of complex, strategic work can only emerge through action, reflection, and more action,⁴⁷ and we agree. Embracing complexity means assigning a greater value to knowledge that comes from a multitude of sources, as we lay out in the following section. Only this kind of deep and continuous learning will allow us to deal effectively with the challenges of complexity.

5. A FULL RANGE OF EVIDENCE IS MOBILIZED IN DECISION-MAKING

The strongest evidence to guide decisions about which investments to fund, implement, and build on and how to improve outcomes will come from combining and integrating what we learn and analyze from multiple sources.^M These multiple sources include: experimental evidence; practice-based evidence; basic and social science research examining causal connections; analyses that combine theory, experience, and other empirical data; and findings culled from the proliferating digital infrastructure. We discuss each in turn.

Experimental Evidence. Experimental evidence from randomized trials and other controlled experiments can be enormously useful and persuasive, as we suggest earlier in this paper. When a

relationships, transactions, forces, and happenstances that occur every day.” Kelly suggests that, given the limited ability to define and control the conditions under which community interventions are implemented, they should prioritize real-time learning, documenting and examining processes, measuring progress toward outcomes, and building capacity to understand and use data. (Kelly, T., Jr. Spring 2010. “Five Simple Rules for Evaluating Complex Community Initiatives.” *Community Investments*, Federal Reserve Bank of San Francisco, 22[1], http://www.frbsf.org/community-development/files/T_Kelly.pdf.)

^M The reason there is no simple answer to the question of what counts as good evidence is that what counts as good evidence “depends on what we want to know, for what purposes, and in what contexts we envisage that evidence being used.” (Nutley, S., Powell, A., and Davies, H. February 2013. “What Counts as Good Evidence?” Provocation paper for the Alliance for Useful Evidence. University of St. Andrews Research Unit for Research Utilisation, <http://www.alliance4usefulevidence.org/assets/What-Counts-as-Good-Evidence-WEB.pdf>).

funder or policymaker wants to know whether a circumscribed, well defined programmatic intervention (be it a program that stands alone or is part of a larger initiative) that has not previously been tried, is achieving specified results for targeted individuals, a randomized trial or other experimental evidence is likely to be appropriate, especially if the question involves possible scale-up of the intervention.

Dr. Donald Berwick, founder of the Institute for Healthcare Improvement and former administrator of the Centers for Medicare and Medicaid Services, describes randomized experiments as “a powerful, perhaps unequaled, research design to explore the efficacy of conceptually neat components of clinical practice...that have a linear, tightly coupled causal relationship to the outcome of interest,” but which, he concludes, “excludes too much of the knowledge and practice that can be harvested from experience.”⁴⁸

We should use experimental methods whenever they are appropriate to the questions we are trying to answer, but we need other methods to answer the questions this method cannot.

According to Anthony Bryk, the evidence we have relied on for the past several decades to shape social policies and programs has come from what he describes as:

*a robust infrastructure... for examining narrow, focused propositions through randomized field trials [that aim] to assess the average effect of some intervention over some non-randomly selected set of sites but provid[e] little or no evidence as to how and why the intervention may work some places and not others.*⁴⁹

Today we are witnessing a deliberate search for methods that “embrace complexity without neglecting rigor and outcomes,” as Darren Walker calls for. We see our work as Friends of Evidence as part of a larger movement to fit the tool to the task by applying a broader array of methods to learn from and improve important outcomes.

Practice-Based Evidence. Practice-Based Evidence is a major contribution toward more evidence-based practice. Lawrence Green, epidemiologist at the UCSF School of Medicine, sees it as a way of getting “a handle on the multiplicity of influences at work in the real world of practice, so that the evidence from our study of interventions and programs can reflect that complex reality rather than mask it.”⁵⁰

Evidence is often narrowly conceptualized as being generated by researchers, evaluators and scientific experts without also including others whose differing relationships to knowledge-generating activities and practices position them to provide singularly valuable and otherwise neglected or missing kinds of evidence. Practice-based evidence takes account of the outcomes valued by participants, the preferences of clients and families, the cultural norms and traditions of diverse communities, and the observations of service providers. **Practice-based evidence creates the generalizable knowledge that emerges when clients on the one hand and expert practitioners on the other are able to examine and reflect together on the processes and results of their work.**

Incorporating practice-based evidence into ongoing work can take many forms, including reflective practice, the blending of surveillance and monitoring, community-based participatory research, and participatory action research. It can add to the likelihood that changes in practice will be well integrated with local strengths and values. And it can increase the effectiveness of interventions.^N In putting

^N For example, studies of vulnerable families have consistently shown that supports and interventions are more

together practice-based evidence, “the real, messy, complicated world is not controlled, but is documented and measured, just as it occurs, ‘warts’ and all.”⁵¹ The Robert Wood Johnson Foundation’s Laura C. Leviton has observed that, despite the complexity of the world it seeks to illuminate, the inclusion of practice-based evidence can be done highly systematically.^o

Basic and Social Science Research on Causal Connections. Information from this type of research is generally underutilized among those most concerned with finding and implementing effective interventions. It includes some of the most exciting new research on human development, including the impact of early stress and other negative psychosocial factors on long-term health and well-being and the importance of children’s early years to brain development.

The implications of this category of information can be considerable. For example, recent medical research produced long-acting reversible contraceptives (LARCs), which have a low failure rate and, once they are in place, do not require the user or her partner to take any further action to prevent pregnancy. In her new book, *Generation Unbound*, Isabel Sawhill argues that, simply by changing the default from opting out of pregnancy to opting in, LARCs could ensure that vastly more children would be wanted children—with all the long-term consequences that shift implies.

Another category of research that should influence intervention design and implementation and resource allocation is the research that sheds light on the importance of the environments where multiple factors interact to create complex problems, such as neighborhoods of concentrated poverty. In a policy brief from *The Future of Children*, Ross Thompson and Ron Haskins report on research that documents how, for children who live in adversity, the stresses of home and neighborhood are exacerbated by poor-quality child care and schools staffed by teachers who are themselves stressed by low income and difficult living circumstances.⁵²

Subtle, hard-to-capture yet critical aspects of interventions, such as the presence or lack of trusting relationships, are hard to document as part of program effectiveness research, but they do come out of other kinds of studies. For example, in 10 years at the Consortium on Chicago Schools Research, Anthony Bryk and colleagues identified the central role of relationships in building effective education communities. They created a metric called “relational trust” to document the social exchanges among students, teachers, parents, and school principals and found that social trust among teachers, parents, and school leaders improves much of the routine work of schools and is a key resource for reform.⁵³

Another example of the gap between the knowledge and practice comes from a recent Institute of Medicine (IOM) report which called attention to the common conditions that underlie so many diverse mental, emotional, and behavioral disorders.⁵⁴ The IOM found that medical and mental health practitioners continue to deal with these conditions as if they were unrelated. In the *American*

effective when they acknowledge and build on client/family values and preferences, and they can be counterproductive or even harmful when they do not. (Attride-Stirling et al., 2001; Watson 2005; Anning et al., 2007; Winkworth et al., 2009; cited in http://ww2.rch.org.au/emplibrary/econnections/Policy_Brief_21_-_Evidence_based_practice_final_web.pdf.)

^o Building generalizable knowledge about practice depends on the ability of expert practitioners to discuss at length the issues they encounter. On that basis, they can develop the implications of their practical experiences. By renewing attention to what program managers and service practitioners know, we can also improve evaluation theory, because their reflective practice can inform both the programmatic and knowledge (external validity) dimensions of theory. See Leviton, L.C. (June 2014). “Generative Insights From the Eleanor Chelimsky Forum on Evaluation Theory and Practice.” *American Journal of Evaluation* 35(2): 244-249.

Psychologist, Biglan and colleagues build on the IOM report to recommend a “drastic shift” toward the widespread adoption of an integrated principle: to increase the prevalence of nurturing environments, and away from a focus on individual problems. They contend that the combination of interventions implied by such a shift would have an enormous payoff in “the reduction of academic failure, crime, mental illness, abuse and neglect, drug addiction and physical illness to levels never before seen in the US.”⁵⁵

A last example of valuable, non-programmatic effectiveness research comes from longitudinal, population level studies that shed light on causal influences by comparing cohorts that encounter different situations at different stages. For example, Princeton University’s Janet Currie looked at the effects on children’s health among those born just before and just after a change in the gasoline lead standards.⁵⁶ Columbia University’s Jeanne Brooks-Gunn used samples of young children in child care in 12 cities and states that had different child care regulations to analyze whether the standards were associated with the quality of care that children actually received.⁵⁷

Analyses Combining Theory, Experience, and Other Empirical Data. Analyses that thoughtfully combined findings from theory, experience, and other empirical data were the source of many great advances in public health during the 20th century, including widespread immunization and the reduction of deaths from cardiovascular disease, stroke, and injuries.⁵⁸ The combined application of theory and empirical data to many leverage points also was behind the spectacular reduction in tobacco use in the US since 1965. Tobacco control efforts in the United States initially focused on individual behavior but evolved to a multi-level, systems-oriented approach that involved industry, legislation, regulations, public health programming, mass media messaging, and the health care system. State by state, it became clear that the complex combinations of strategies were more effective than any single intervention. The two states that undertook the most comprehensive interventions (California and Massachusetts) doubled and then tripled their annual rate of decline in tobacco consumption relative to the other 48 states. These two states were able to achieve change at scale largely because they acted on every kind of intervention that seemed to hold promise.⁵⁹

This kind of broad-ranging analysis has occurred much less frequently in the human services arena. University of Pennsylvania family sociologist Frank Furstenberg writes that the separation of non-experimental research (showing the powerful influence of the family) from experimental research (designed to apply the lessons of that literature) has led people to position macro- and micro-level policies as strict alternatives. He contends that we would do much better by integrating and acting on knowledge from both arenas.⁶⁰

In addition, we must find systematic ways to include the indigenous wisdom of those who have the most accurate, precise and detailed understanding of local communities and circumstances. Their contribution to understanding the problems that interventions aim to address, existing strengths and resources that can be built upon, and the nature of effective solutions is increasingly seen as essential.

The Proliferating Digital Infrastructure. Digital infrastructure is making possible the use of innovative data collection methods, supplementing traditional evaluation methods such as surveys and interviews. Data aggregation at a scale previously unattainable—crowd-sourced data, broadly inclusive stakeholder engagement, problem identification and problem solving, GPS mapping and tracking tools, highly sophisticated data visualization techniques, and infographics—all can deepen our understanding of complex social phenomena, help us see trends and gaps, and inform solutions. Advances in technology are also making it possible to understand emergent phenomena within complex systems and networks,

including the role of neural network connectivity in learning. And the new digital infrastructure allows diverse groups of people in multiple locations to generate, collect, and use data—often involving no evaluators at all and thus minimizing the distinction between information generators and users.⁶¹ The development of the capacity to take full advantage of these new ways of working with data, could, according to Karen Pittman of the Forum for Youth Investment, be seen as an intervention in and of itself.

Indeed, it is hard to predict how the aggregation of data and data users in crowd-sourcing and social/political movements will affect our ability to understand and improve our interventions.

The practice of learning from a full range of sources, such as those we have suggested here, trades some certainty for approximations. This is a choice that many of the best evaluators are already making, according to Jeanne Brooks-Gunn. Such trade-offs can be valuable. Douglas Nelson, president emeritus of the Annie E. Casey Foundation, believes that “interventions that have the potential to change things at scale have so many dynamic elements that there’s got to be a way to find an approximation of the elements that tend, when they’re present, to predict a result.” He has concluded that, “If you don’t settle for a *tendency* to produce results, you end up with an intervention that is too small, too unreal, too artificial. You force people to find insufficient answers or to solve a problem other than what we’re worried about.”⁶²

We would thus make room for judgment, thoroughly infused by evidence, as an essential guide to action. As Stanford University Graduate School of Business Professor Susanna Loeb puts it, “Even with the best available information about the present and the past, decisions about the future rely on judgments as well as knowledge.”⁶³

IV. OPERATIONALIZING A POWERFUL APPROACH TO EVIDENCE

We believe the time is right for fresh approaches to generating and applying evidence of social interventions’ effectiveness. A new “evidence mindset” already is becoming visible as system-reform and community-change efforts increasingly base their interventions on a variety of sources of knowledge; creatively deploy results-based approaches, continuous-improvement methods, and networked learning communities to extract knowledge; and use what they are learning to improve interventions and to achieve results at greater scale.

In this section, we propose an initial framework to generating and using evidence. None of the individual elements of our framework is new; all have been identified and examined in existing bodies of knowledge (including implementation science, continuous quality improvement, performance management, action research, Results-Based Accountability,[™] and other fields), and we draw on those sources here. What is rarely done, however—and thus offers new opportunities—is to intentionally and systematically create the conditions under which these diverse and potentially synergistic ways of generating knowledge can be coherently integrated, implemented, rigorously examined, and sustained over time. This approach has been described as a “science of improvement” and as “a productive synthesis” that melds “the conceptual and methodological strength associated with experimental studies to the contextual specificity, deep clinical insight, and practical orientation characteristic of action research.”⁶⁴

By linking these elements as parts of an overall new approach, we hope to further encourage system-reform and community-change efforts to use the combination of methods that:

- Inform the initial design of interventions most thoroughly, so that each new effort builds on the best foundation of prior knowledge; and
- Can be best adapted to the enterprise at hand, to produce (a) the information needed to guide on-going implementation and (b) data about what worked, for whom, and why as well as what didn't work—that is, where an intervention falls on a “spectrum” of effectiveness.

The elements we see as most important for this approach are shown in Fig. 3. We then describe each in more detail, with illustrations drawn from current complex system-reform and community-change initiatives.

Even a quick scan of Figure 3 reveals that the elements of a comprehensive approach are more likely to be effective when they are linked together. In fact, many initiatives use many, or in some cases all, of these elements as part of their learning agenda or their continuous improvement plans. Here, we focus on each in turn because it is useful to understand the characteristics of each element as we consider how they function together to provide a more comprehensive approach to evidence.

**Fig. 3:
ELEMENTS OF A POWERFUL APPROACH
TO EVIDENCE**

1. Many sources of evidence inform intervention design.
2. Results shape implementation.
3. Goal-oriented networks accelerate knowledge development and dissemination.
4. Multiple evaluation methods fit diverse purposes.
5. A strong infrastructure supports continuous learning for improvement over time.

ELEMENT 1: MANY SOURCES OF EVIDENCE INFORM INTERVENTION DESIGN

Designers of new interventions, strategies, system changes, and policies look first to what has been done in the past and what is known about the effects of past initiatives. The typical approach, whether in government agencies or foundations, at the federal or local level, is to begin with various lists of evidence-based programs, or EBPs. Although their criteria for inclusion vary, these sources identify programs that have shown significant effects as measured by randomized-control trials or other experimental or quasi-experimental methods. Although the number and types of programs that have been established as evidence-based is limited, many system or community leaders who want to improve results for sizable or diverse populations find them to be a reasonable starting point.

We recommend that planners draw on knowledge from other sources, too. Useful sources, some of which are described in Section III of this report, include: comparative, developmental and formative evaluations; studies of implementation and dissemination; research on specific factors and contexts; knowledge from practice and lived experience; basic and social science research; analyses that combine theory, experience, and empirical data (including epidemiological and econometric research); and digital surveillance systems that collect and compare data on national, state, and local innovations over time and across jurisdictions.

Some skilled initiative planners have figured out how to combine evidence from multiple sources in creative ways. The New Haven (CT) MOMS Partnership, a community-based initiative that addresses maternal depression (Fig. 4), and the Promise Neighborhoods initiative in Minneapolis (Fig. 5) are good examples. Leaders of both programs:

- **Understood that no one source of evidence would be sufficient** to design the innovative strategies needed to achieve desired results;
- **Investigated EBPs** with an appreciation for what they could provide and with an understanding that programs alone are unlikely to be able to meet the full spectrum of community needs;
- **Drew from other sources of evidence** to develop their theory of change and their specific interventions;
- **Devoted time and resources to tapping the knowledge of parents, youth and residents**, confident that this would shape interventions and strategies in practical and unique ways; and
- **Built in a process for distilling new knowledge and incorporating it** into strategies, reflecting a belief that the process of design and creation is on-going, not static.

Fig. 4:
THE MOMS PARTNERSHIP

MOMS was designed by clinicians and researchers at the Yale School of Medicine to address widespread maternal depression in under-resourced and overburdened New Haven neighborhoods and to test the efficacy of a new blend of clinical practice and community-based strategies. To develop MOMS' theory and design, the team drew on multiple sources of evidence.

They began with an understanding of maternal depression based on years of basic science, epidemiological and clinical research (especially during the last 15 years, as knowledge of physiological and neurological contributors has grown) and on recent advances in understanding the impact of stress. Next, they studied evidence about the effectiveness of tested interventions, especially relevant evidence-based programs. From all of this information, researchers extracted what they could and adapted the knowledge into new approaches that might be tested.

With this information, the team set out to understand the lived experience and perspectives of low-income mothers in New Haven. Over 8 months, they met with approximately 1,000 mothers to learn about their experiences, their perceptions of stress, the barriers they faced in providing for their children, the impact of community safety and poverty on their lives, and their expectations for the future. The picture that emerged pointed to solutions that no other source of information had indicated. For example, lack of diapers for young children emerged as a major source of stress for the mothers, so providing diapers became a component of the new intervention. Subsequent data collection proved this to be one of the most influential factors in helping mothers reduce stress in their lives.

After launching the intervention, the MOMS project team learned more from emerging neuroscience about the impact of parental actions on a child's development of executive function skills, even during the earliest years of life. Responding to this new knowledge, the MOMS Partnership is developing an additional component to strengthen mothers' executive function, thus increasing the likelihood they will be able to help their children build the same capacity, while simultaneously helping them move into employment, since research indicates that the skills developed in employment preparation are the same or similar to executive function skills.

The MOMS project and research team was well-connected to resources that could provide them with the most recent findings from brain science; they also had resources to conduct extensive interviews with parents. If they had not, and if they had relied only on programmatic evidence, the intervention they designed would have been less robust, less attuned to the realities of mothers' lives, and perhaps less likely to succeed.

Preliminary data from the MOMS Partnership are encouraging. The mothers served thus far show statistically significant reductions in parenting stress and depression symptoms and equally significant increases in positive parenting practices and executive functioning. Additionally, program participation continues to be strong: adherence to 8 sessions of a skill building stress management course range between 92-96% for mothers who have enrolled.

**Fig. 5:
PROMISE NEIGHBORHOODS IN MINNEAPOLIS**

Situated on the city’s north side, with the nonprofit **Northside Achievement Zone (NAZ)** as lead agency, this initiative had to align its design with the U.S. Department of Education’s desire to anchor *Promise Neighborhoods* in the best available evidence. The Department required sites to deploy evidence-based programs to create a “pipeline” of cradle-to-career supports. NAZ, like other successful applicants for *Promise Neighborhoods* implementation grants, investigated and chose a number of EBPs as pipeline components. NAZ used the search for relevant EBPs as a first step in initiative design, and then augmented it with a well-defined process for adapting interventions and developing new “solutions” (as *Promise Neighborhoods* interventions are often called) that are tailored to the local area, a predominantly African-American community with considerable resident assets but also high poverty and low rates of school success.

NAZ was operating as a pilot program working to mobilize parents in new ways and to layer and combine multiple interventions and customize supports around the needs of each child and family. However, in order to attend to population-level results and improve effectiveness of existing solutions, NAZ leaders needed to adopt and embed the learning from evidence-based programs across their continuum of solutions – but often in new ways.

To do this, NAZ leaders invented a process they call the “NAZ Seal of Effectiveness.” It aims to build on the best knowledge available by adapting existing models or creating new solutions. A panel of local leaders, residents, researchers, and program experts, augmented by national consultants in the subject area, synthesize all they know from research and experience into an intervention that NAZ and its partners will put into practice. The essential ingredients are specified, along with indicators that will show whether the ingredients are used appropriately. Implementation is carefully tracked to assess evidence of impact as well as fidelity to essential ingredients. A NAZ “Results Roundtable” meets regularly, using assessment data to determine if the intervention is being implemented as intended, having the desired effect, or needs to be adapted to increase the chances of success. NAZ is establishing roundtables for several components of the cradle-to-career pipeline, with the goal of eventually having this rigorous and adaptive oversight for each segment of the service continuum.

ELEMENT 2: RESULTS SHAPE IMPLEMENTATION

Leaders of change initiatives systematically review implementation in relation to results, and then adapt, improve, refine, and/or redesign interventions. They develop solutions for complex problems by creating an on-going cycle of problem solving, strategy testing, observing what’s working, for whom and how, as well as who’s not well-served by an intervention; continuing the strategies that are effective and adapting or dropping those that aren’t, and building on success to achieve solutions at greater scale. This process of analysis and action is distinguished by: a relentless focus on results; powerful use of relevant data; a commitment to spending time understanding implementation and linking performance to outcomes; and the management discipline to repeat this process continuously, in a way that engages and motivates the people who participate in it.

Examples of this type of process are emerging from many different domains and traditions, accompanied by a variety of tools. The most frequently cited are results-based management, performance management, continuous quality improvement, and the techniques collectively known as implementation science.

These bodies of knowledge and experience have several common characteristics. We emphasize three that seem to affect the success or failure of on-the-ground implementation—that is, their ability to develop strong enough data about performance and results over time to generate evidence that can guide others’ similar efforts. They are:

- **Clear definition of desired results, with results defined as improvement in the conditions or status of people or neighborhoods.** All of these approaches start with the end in mind, and the work they do is oriented toward achieving specific results and improving the indicators of those results.
- **A commitment to collecting, generating, analyzing, and using relevant data.** These approaches depend on data that are: relevant, generated by reliable measures of the desired results, regularly available for management and accountability purposes, and viewed as important by all who use them.
- **A management and organizational culture that orients to results, values data, seeks out errors and failures as unique opportunities for learning and improvement, and creates time for reflective practice** based on feedback and learning. More easily aspired to than realized, the continual process of feedback and learning described here creates a “culture of results” that becomes normal practice for an organization, a partnership of multiple collaborators, or a large system (e.g., public agency).

These characteristics support a rigorous process by which organizations, initiatives, or systems can start from the best available knowledge and then adapt and improve to achieve better results over time. Instead of implementing a single program model that may not succeed in all places for all people, this process intentionally responds to multiple interactive forces embedded in local context, population shifts, and changing needs—with rigorous documentation of both the adaptations and their effectiveness. This approach also attends carefully to the details of implementation, realizing that a failure to execute or implement as intended may be the reason why a proposed solution does not succeed (and recognizing that there are other factors that can limit success).

Putting this approach into practice requires well-developed tools. We mention several here to highlight important work that has been done and to underscore that not every initiative has to invent tools from scratch. The tools and processes include:

- **Results-Based Accountability™ (RBA).** Developed by Mark Friedman of the Fiscal Policy Studies Institute, this approach to results-based thinking and management provides a disciplined construct and vocabulary for establishing (a) desired population quality of life results (e.g. Healthy People, Safe Communities, Clean Environment), (b) indicators for measuring quality of life results, and (c) a process for developing and tracking the effectiveness of solutions that “turn the curve” on the indicator baselines for each result. A major contribution of RBA has been to distinguish Population Accountability for conditions of well-being for an entire population in a

geographic area (e.g. neighborhood, city, county, state or nation) from Performance Accountability for the quality of service delivery to program customers. RBA is proving useful in community initiatives such as *Promise Neighborhoods* and in the government and nonprofit sectors in the United States and at least 15 other countries. In addition to the tools developed by Friedman, the Results Leadership Group and others, an important dimension of implementing RBA has been adopted by the Annie E. Casey Foundation's Results-Based Leadership program, which unites the use of RBA with intensive training in creating results-based cultures.

- **Performance management.** The movement to adopt this approach, which is increasingly becoming a central tenet of governmental and non-profit organizations, is motivated by many of the same purposes that underlie RBA. Performance management advocates and provides tools for a careful process of setting performance measures and then regularly tracking the effects of implementation against them.
- **Implementation science.** This growing body of work calls attention to the fact that what is done and how it is done often determines whether the desired results are achieved and, in addition, helps identify unintended effects and results. Implementation research has identified a “decision support data system”⁶⁵ as essential to the processes of establishing innovation, supporting practitioners, and assessing outcomes. Implementation science also emphasizes the importance of distinguishing between adaptive issues and technical problems and using systematic improvement cycles, such as the “Plan-Do-Study-Act” approach adapted from private industry. The National Implementation Research Network (NIRN) has been a leader in this effort, defining a set of “active implementation drivers” to guide public systems and non-profit organizations.⁶⁶

Three examples illustrate the different ways that initiatives or organizations trying to solve complex problems use the feedback cycle and a culture of learning for results to evolve effective solutions and improve results for children, families or communities. The Northside Achievement Zone (NAZ, Fig. 6) found it essential to use a well-crafted tool such as Results-Based Accountability™ to help develop its process. The Cincinnati Children's Hospital Medical Center (Fig. 7) has been building a culture of innovation and results since 2002, and its experience illustrates how a large hospital—a complex system in itself—uses the development of a culture of results to achieve impact and develop evidence. And the national non-profit known as LIFT (Fig. 8), which organizes volunteers to help people move out of poverty, illustrates the value of looking for non-traditional forms of data as a way to shape innovations and document their effects.

Fig. 6:
RESULTS SHAPE IMPLEMENTATION

The **Northside Achievement Zone (NAZ)** uses a steady flow of data, careful day-to-day management reviews, and periodic reflection on results to establish a cycle of rapid, continuous learning and refinement in its efforts to promote “cradle-to-career” success.

The approach begins with 10 results and 15 indicators specified by the U.S. Department of Education, which all *Promise Neighborhoods* grantees pursue. NAZ has gone a step farther, however, and structured its entire approach around a Results-Based Accountability™ process. To support the process, NAZ utilizes a hosted data system, NAZ Connect, which collects detailed data on the implementation of each strategy and its impact at multiple levels: on individual children, cohorts of children, and all children in the neighborhood population. NAZ and partner staff review the data frequently to identify children who are falling behind and those for whom solutions are not working. The Seal of Effectiveness process (see Fig. 5) is used to consider what needs to change so that local interventions will help the children who are not yet succeeding. This data-based process is in turn embedded within an organizational culture that encourages looking at errors as well as successes; encourages innovation; and engages and learns from participants whose voices and perspectives are invaluable but often neglected.

Preliminary data indicate that the early childhood interventions for which this process was first used are working. 2013-14 school year data revealed that school readiness increased by 6% “zone wide,” as indicated by the Beginning Kindergarten Assessment.

Fig. 7:

FEEDBACK FROM FAMILIES LEADS TO SOLUTIONS

Cincinnati Children’s Hospital Medical Center (CCHMC) ranked third among 183 hospitals surveyed in *U.S. News and World Report’s 2014-15 Best Children’s Hospitals* ranking. Leaders credit this status to: (1) an ongoing focus on improving patient outcomes (always a primary result); (2) a careful and institution-wide use of data to develop and refine solutions that respond to context and to feedback from patients, even when it means crossing the boundaries between areas of work; and (3) creation of an organizational culture committed to constant self-reflection and improvement. Hospital administrators say the culture of results has led to more creative solutions, which have become institutionalized as data reinforce their effectiveness.

An early example of this process involved the hospital’s efforts to reduce hospitalizations of children with asthma. When a team of health care professionals and hospital administrators examined data on the children admitted for asthma, they found that many came from the same neighborhood and the same apartment buildings, and they often had the same landlord. They also learned that families often didn’t pick up their children’s medications, which was a barrier to asthma care.

To fix the problem, a CCHMC official says, “We couldn’t just think and act within hospitals. We needed to go to the source of the issues.” That meant talking to the parents of children with asthma and to community partners that interact with the children, such as the child’s school nurse.

The conversations revealed a problem with mold in the families’ homes and difficulty accessing needed medications. So CCHMC developed a layered strategy to address the root causes of the problems. The hospital engaged Legal Aid to help families put legal pressure on landlords to remove the mold and ensure their homes were healthy environments—even putting a Legal Aid office in the hospital’s clinic. To make it easier to pick up medications, CCH first tried making them available at the clinic. When that still proved ineffective for many families, the hospital arranged for mobile delivery of medications.

At each step, the hospital’s asthma results team examined which strategies worked for which families and recognized that some families required different interventions. The interventions were then adapted, with progress against defined performance targets continually monitored. The customized interventions were successful. Asthma readmissions to CCHMC for children in the county decreased from 7.2 admissions per 10,000 children in 2009 to 4.7 admissions per 10,000 in 2014.

**Fig. 8:
DATA HELP TO PREDICT SUCCESS**

LIFT operates in six cities to offer its clients (called members) a flexible, customized package of services in a community setting. At the heart of LIFT's approach is what its founder, Kirsten Lodal, describes as a simple idea: that all people, including people in poverty, are experts in their own lives and deserve to be the architects of their own success. For LIFT, that means recognizing that each member they serve is a whole person with a holistic set of needs, strengths and aspirations, and LIFT's work with each member must be defined by the goals and priorities that are most important to them. Lodal and her team were uncomfortable that LIFT had no way to systematically listen to the thousands of people they served each year, and had failed to ask the members themselves how well LIFT understood and met their needs. So they co-developed a measurement system (in partnership with Keystone Consulting) that is designed to capture evidence through a method called "Constituent Voice."

LIFT conducts a micro-survey (4-6 questions) after every interaction with a member (i.e., an individual or family participating in services), with rotating questions that address things like relationship quality and service importance. The survey is administered using simple, touch-based technology on iPads, which appeals to members with literacy challenges or discomfort computers. LIFT regularly analyzes the data to help its team understand which interventions are having what effect and for whom. Feedback is incorporated into key program strategy, design, and delivery decisions. And every other week, staff review and discuss the results. "We continuously close the loop," Lodal says. "We take the results back to our members to further make sense of the data and to ask more questions. We brainstorm directly with members to uncover ways LIFT can improve how we support them."

As LIFT has examined its constituent voice data against its objective measures of economic progress, it is finding interesting relationships. Members who give LIFT higher Constituent Voice scores achieve three times as much progress toward their goals as detractors (or low-scoring members). In looking for indicative metrics, LIFT has found that from among the 20 questions tested in the surveys, the question that correlates most closely with economic progress is not service quality or service relevance; it is "social connectedness" to both the community and community resources. Further, the top indicators of success are related not just to a member's social connectedness but also to LIFT's quality of service and the quality of relationship with LIFT. Lodal's conclusion is that, "For anti-poverty organizations, this speaks so forcefully to the importance of getting better at doing things like bolstering supportive relationships and building social capital if we ultimately want to see people's financial security grow."

While the use of this data as part of LIFT's feedback and innovation cycle is still new, it provides for the first time a data-based way to test the organization's operating assumptions. It becomes an important component of a multi-faceted evaluation, offering the possibility of a predictive indicator. "Instead of waiting 20 years for the results of a longitudinal study, we can ask questions now that can reliably indicate future success," Lodal says. "As David Bonbright, whom we're partnering with, says, 'Who needs causality when you have predictability?'"

ELEMENT 3: NETWORKS ACCELERATE KNOWLEDGE DEVELOPMENT AND DISSEMINATION

Goal-oriented networks are proving to be a useful tool in the quest to accelerate learning about what works, for whom, and why. As individual communities, initiatives, or systems work to improve and to achieve better results, being in a networked relationship with others who share the same goals can multiply and amplify learning.

The goal of ensuring that reform efforts learn from one another is not new, of course. In almost every multi-site initiative, funders have supported cross-site meetings where participants exchange ideas, learning communities in which leaders analyze their successes and failures, and technical assistance in which people engaged in similar work learn from one another.

The concept of networked learning goes several steps further. It involves creating a cohort or network of initiatives that, in the interest of shared learning, are willing to align or standardize elements of their design and implementation to allow for more powerful comparisons across sites. . Karen Pittman points out that the Forum for Youth Investment, which she heads, has found that learning networks can also be powerful within the same community, especially when “undergirded by common outcomes, frames, and measures.” Networks are also strengthened when members are given time to reflect on their experiences together. Intensive reflection, learning, and adaptation of solutions or strategies are hard-wired into all aspects of the network’s design and operation. Done well, this approach also sets specific deadlines to arrive at particular goals, intensifying focus and bringing discipline and accountability to the approach

For example, the Carnegie Foundation for the Advancement of Teaching is investing in networked learning to test its effectiveness as a way to accelerate and enrich the evidence available to improve schools, classrooms, and the education enterprise in general. According to President Tony Bryk:⁶⁷

It is now possible to accumulate evidence network-wide because of the use of common frameworks for defining problems and hypothesizing solutions along with common measures for examining outcomes. The breadth of information generated across contexts and participants enhances the possibilities for innovation and expands insights beyond those arising in any one place.^P Thus, improvement research structures systematic inter-organizational learning.

Networked learning requires careful conceptualization of how problems will be characterized and addressed, solutions adapted, and information assessed for reliability and transferability from one network member to another. To establish an appropriate environment for learning, the Carnegie Foundation has developed the following core principles:⁶⁸

- 1. Make the work problem-specific and user-centered.** Inquiry starts with a single question: “What specifically is the problem we are trying to solve?” It then enlivens a co-development orientation, engaging key participants early and often.

^P The Foundation contrasts networked learning with a reliance on proven programs, which are expected to stay static and be “replicated.”

2. **Variation in performance is the core problem to address.** The critical issue is not what works but rather *what works, for whom, and under what set of conditions*. Aim to advance efficacy reliably at scale.
3. **See the system that produces the current outcomes.** It is hard to improve what you do not fully understand. Place special emphasis on going to see how local conditions shape work processes and engaging a broad range of constituents who can illuminate these processes. Make your hypotheses for change public and clear.
4. **We cannot improve at scale what we cannot measure.** Embed measures of key outcomes and processes to be tracked, to determine if change is an improvement. We intervene in complex organizations, so anticipate unintended consequences and measure these, too.
5. **Anchor the improvement of practices in disciplined inquiry.** Engage rapid cycles of Plan, Do, Study, Act to learn fast, fail fast, and improve quickly. That failures may occur is not the problem; that we fail to learn from them is.
6. **Accelerate improvements through networked communities.** Embrace the wisdom of crowds and leverage the divergence of their experience. We can accomplish more together than even the best of us can accomplish alone.

One of the networks supported by the Carnegie Foundation (Fig. 9) illustrates how these principles structure a learning network in practice. As is clear from Carnegie’s work, accelerating the development of learning through a network is one element of a more comprehensive approach to learning—one that complements many of the principles and practices described in earlier sections of this report. When the complete package is put in place, however, the existence of a network that uses “the wisdom of crowds,” as Carnegie puts it, can both multiply learning and increase scale dramatically.

**Fig. 9:
GOAL-ORIENTED NETWORKS ACCELERATE LEARNING AND RESULTS**

Community College Pathways began in July 2010 as a Networked Improvement Community of 27 community colleges and three major universities tackling the problem of low student success in developmental mathematics in higher education. The network has since expanded to include 51 institutions across 15 states serving nearly 9,000 students placed into developmental mathematics.

With support from the Carnegie Foundation for the Advancement of Teaching, network members have (a) developed new curricula to support ambitious learning goals around “mathematics that matters”; (b) taught students using new research-based pedagogy; and (c) tested and implemented strategies to help students persist in college. To promote consistently high performance as the initiative grows, and to ensure effective implementation across contexts, the networks funded by Carnegie use continuous improvement strategies.

The results have been overwhelmingly positive; Pathways students in developmental mathematics have tripled the success rate in half the time compared to outcomes prior to the establishment of the network.

ELEMENT 4: MULTIPLE EVALUATION METHODS FIT DIVERSE PURPOSES

The examples described thus far illustrate the benefits of generating knowledge by using continuous improvement processes to achieve results. These techniques are not added on to an initiative's implementation; they are part of the implementation and inherent to the way complex change processes are designed and managed.

In addition, these initiatives (and others like them) make use of a variety of evaluation techniques to gain a more complete understanding of the impact of specific parts of their intervention, as well as of the whole. In fact, the combination of continuous improvement processes and more traditional evaluation activities, well-integrated, can become part of a larger tool box for generating evidence.

Many complex initiatives now use a range of evaluative approaches to illuminate different aspects of an initiative's effect. These include: (1) formative evaluations, to help understand initiatives during their early stages and to identify when implementation has reached the point at which an impact evaluation would be useful; (2) developmental evaluation, an approach particularly attuned to evaluating complex initiatives, focusing on what is being developed and to what effect through innovative engagement with program leaders and, when appropriate, program participants; (3) case studies, which also examine the processes of implementation and capture a subtler range of information than answers to the question, "Did it work?" by going on to ask, "How well did it work? For whom? Did it work as we expected?"; (4) new, rapid forms of randomized-control trials, which can provide information on the impact of specific programmatic interventions for specific populations; and (5) various forms of comparison group evaluations that, for example, improve understanding of how an initiative's impact compares to its absence in a similar community or system.

One advantage of multiple methodologies is that the combined information they produce creates a deeper understanding not only of what is working but why – which in turn allows that knowledge to be more readily applied in other contexts. Three examples (Fig. 10) illustrate how complex interventions are being evaluated through multiple means. In each, it is the combination of continuous learning through core management processes, plus critical data from multi-faceted evaluation, plus engaging and activating providers and participants in the process of improvement, that provides the information leaders need to steer the initiative, adapt it for greater success, and build a knowledge base about implementation and impact that will help the broader field.

Fig. 10:
USING MULTIPLE METHODS TO EVALUATE COMPLEX INITIATIVES

Cincinnati Children’s Hospital Medical Center (CCHMC) conducts research and evaluation on the effectiveness of its health care service delivery and its impact on community health. The most frequent use of evaluation (in addition to the on-going use of rapid-cycle improvement, tracked through management practices) is for new interventions. These are often formally evaluated, using the hospital’s own resources or federal or foundation grants, through clinical trials (both longitudinal studies with large cohorts and short-term RCTs). In addition, CCHMC often uses process documentation and action research to gather information about community impact and/or community reaction to new or on-going service delivery. The hospital’s strategic use of multiple methodologies is coordinated in some instances by researchers employed by CCHMC and often involves partnerships with outside evaluators and/or academic institutions.

The **CareerAdvance® Program** of the Community Action Project (CAP) of Tulsa, Oklahoma, is based on the belief that family economic success will protect and enhance gains made through high-quality early childhood programs. The program provides sector-based workforce development to adults in low-income families whose children attend Head Start and Early Head Start. Key elements include training courses, peer mentoring and coaching, tuition payments, performance incentives, tutoring, and wrap-around services. Evaluators from Northwestern University and the University of Texas are conducting formative studies to understand how and why the program is having effects. The evaluators’ performance data on specific components help staff and administrators adapt program components and procedures to address emerging issues and improve results. As CareerAdvance® grows, the evaluation methods will include comparisons with a sample of CAP parents and families whose children are in the program but the parents are not in CareerAdvance.® In addition, CAP has just started an RCT for parents who are going into developmental education or English as a Second Language (ESL) instruction. Complementing these components of the evaluation, researchers also are documenting the program through case studies on the process of learning and continuous improvement and on the braiding together of interventions.

The **Northside Achievement Zone (NAZ)** is using a randomized control trial, conducted by the University of Minnesota, to evaluate the effectiveness of the Family Academy College Bound Babies curriculum. The RCT, which uses families from the program’s waiting list as the comparison group, is designed to produce data to provide NAZ with the needed feedback about program impact, while ensuring that all families can receive the program. However, this RCT is only one component of NAZ’s overall research and evaluation process. In partnership with the university, NAZ’s primary focus is conducting process evaluations of the overall development and implementation of collective impact solutions.

ELEMENT 5: A STRONG INFRASTRUCTURE SUPPORTS CONTINUOUS LEARNING FOR IMPROVEMENT

Initiatives that involve multiple modes of learning and adapting require a strong infrastructure, including organizational capacity, human resources, and management processes. The initiatives we have described each created infrastructure in a different way based on their organizational context, the resources available, and managers' leadership styles. But all were able to generate evidence that enable them to apply processes rigorously and systematically, again and again, over time. Looking across these and other initiatives, we see the following essential ingredients of a strong infrastructure that supports learning and results:

- **Leaders that prioritize results and continuous learning**, making an explicit commitment to a culture of results and to continuous learning. The entire enterprise can be built on this approach (as with NAZ's Promise Neighborhoods initiative), or it may occur as implementers seek to improve service effectiveness within one system (as with CCHMC).
- **A well-defined, continuous, and highly valued process for reviewing data**, incorporating feedback, and adapting or changing interventions. One striking thing about the initiatives described in this paper is how clearly articulated and carefully maintained their feedback cycles are.
- **Data systems that provide information on both population-level impact and individual effects.** The data systems of the initiatives described here are constantly evolving to better support learning and accountability. The resources available for these systems vary widely, but each devotes a substantial part of organizational or initiative resources to data systems and processes, viewing them as essential to every aspect of their operations.
- **Program staff who support continuous learning** and are trained in it. The staff capacity and skills necessary to do this work—including an interest in gaining a true understanding of the causal connections and the extent to which efforts are achieving desired results, enthusiasm for continuous program improvement, and dissatisfaction with business as usual—often must be acquired. Leaders must pay special attention to how staff are recruited, trained, evaluated, and supported.
- **Strong community support for the initiative and the involvement of consumers, residents, and/or parents in the process of continuous learning.** NAZ, the MOMS Partnership, and LIFT are particularly clear examples of how parents, community residents, and initiative members assume leadership roles in the learning process. Moreover, all of these initiatives do more than simply “engage” constituents.
- **Staff and/or partners with clear responsibility for maintaining the learning process.** This type of work cannot simply be added to staff's other responsibilities. It requires dedicated time by staff, consultant, or contracted resources. The rigor and the on-going nature of these processes cannot be sustained without dedicated time for data development and analysis, managing the learning process, producing reports that engage partners, and so on.

- **Resources to support continuous communication among collaborators, to support time and training needed for collaboration and for follow-up to collaboration.** These approaches require the on-going, active engagement of many partners, and resources are necessary for these processes of collaboration, or collective impact, to be effective and sustained.
- **A relationship with one or more research or academic institutions or outside evaluators.** This somewhat unexpected element is present for most of the organizations described here. In all of these efforts, the academic/community partnerships or the relationships with other outside evaluators establish a very clear and productive understanding of respective roles and responsibilities.

It is challenging for organizations that implement complex initiatives to develop and sustain funding for the kind of infrastructure described above. Most of the initiatives described here have federal or foundation grants that support infrastructure and allow intensive, results-based processes and learning to take place. As a number of funders have recognized, these resources cannot continue to be viewed as “special.” **Infrastructure for learning and improvement is foundational and necessary for effective, result-based work, and building this type of capacity is inseparable from the process of achieving results at scale.** This presents a challenge to initiative managers and funders alike. Funding this infrastructure and capacity may not lend itself to the traditional demands for sustainability, which often focus on programmatic elements, even though it may be precisely the infrastructure for the results-based, continuous improvement way of work that produces better results, more cost-efficient interventions and results that create savings over time.

V. NEXT STEPS

If the multi-dimensional framework for evidence described in this paper is to become a routine approach to creating knowledge to solve complex problems—the “new normal”—current partners will need to adopt new perspectives and roles and many things would ultimately change. For instance:

- **Public and private funders** would ensure that knowledge development and dissemination accelerates the achievement of better outcomes at scale, by investing in complex interventions and generating evidence from them in multiple and often complex ways. This may involve facing the fact that these initiatives address problems for which we actually do not already have solutions, and that as solutions are developed they may not be lasting – since they occur in complex systems whose moving parts and interactions are constantly changing. Foundations and other private-sector funders would: pay more attention to units of analysis that are larger than individual programs; actively encourage the use of multiple sources of evidence; require the engagement of multiple stakeholders in the interpretation and ideally in the collection and analysis of evidence; and help to disseminate evidence from these approaches across the field, with careful attention to the effects of dissemination (i.e., determining if what is disseminated is actually used in the field).

- **A range of institutions** would help create the conditions for developing, testing, and improving solutions to “adaptive” problems. They would develop capacity to work with and disseminate measurement, coordination, communication, and data-sharing tools that could strengthen efforts by public and nonprofit agencies, communities, states, and the federal government to improve outcomes. For example, they could: support the creation of “networked improvement communities,” as described in this paper; provide incentives and tools for using continuous improvement processes to change systems in multiple states or institutions; invest in developing useful indicators and intermediate markers of hard-to-measure outcomes; and engage with partners in ways that build and sustain a culture of improvement.
- **Academic institutions and peer-reviewed publications** would broaden their views on evidence and replace the current hierarchy of evidence with a taxonomy of evidence that aligns research methods with the questions and challenges they are best suited for. There are significant advances on this front, as leaders in various academic fields call for research that is more relevant for decision making and improved outcomes.

The Symposium on the Future of Evidence will explore possible next steps by these and other entities to support a broader approach to evidence. Ideas that emerge from the symposium will form the beginning of work by the Friends of Evidence to chart the most feasible and productive steps and to collaborate with the federal government, foundations, and others as they move in this direction.

As we look ahead, we are impressed by the many individuals and organizations throughout the country that are struggling with the challenge of generating, applying, and disseminating knowledge that is strong enough, deep enough and relevant enough to be a credible guide to more effective interventions and—ultimately—to substantially improved outcomes. Leaders of several initiatives will describe some of their approaches to evidence at the Symposium, some of which we portray in this paper. These examples make clear that our vision of broader approaches to evidence is not a pipedream. It is happening in places across the country. **The leaders of these initiatives are demonstrating that their new approaches to evidence will simultaneously lead to clearer accountability, more-effective initiatives, and accelerated learning about how to improve outcomes at scale.**



ENDNOTES

- ¹ <http://tcf.org/blog/detail/concentration-of-poverty-an-update>
- ² Shonkoff, J.P. (January/February 2010). "Building a New Biodevelopmental Framework to Guide the Future of Early Childhood Policy." *Child Development* 81(1): 357–367.
- ³ Sawhill, I.V. and Karpilow, Q. (July 2014). "How Much Could We Improve Children's Life Chances by Intervening Early and Often?" CCF Brief # 54. Washington, DC: Brookings Institution Center on Children and Families.
- ⁴ Sawhill, I.V. and Karpilow, Q. (July 2014). "How Much Could We Improve Children's Life Chances by Intervening Early and Often?" CCF Brief # 54. Washington, DC: Brookings Institution Center on Children and Families.
- ⁵ Coalition for Evidence-Based Policy, (July 2013). "Demonstrating How Low-Cost Randomized Controlled Trials Can Drive Effective Social Spending: Project Overview and Request for Proposals."
- ⁶ Coalition for Evidence-Based Policy, (July 2013). "Demonstrating How Low-Cost Randomized Controlled Trials Can Drive Effective Social Spending: Project Overview and Request for Proposals."
- ⁷ Shonkoff, J.P. (January/February 2010). "Building a New Biodevelopmental Framework to Guide the Future of Early Childhood Policy." *Child Development* 81(1): 357–367.
- ⁸ National Scientific Council on the Developing Child. (2004; updated October 2009). "Young Children Develop in an Environment of Relationships." Working Paper No. 1, http://developingchild.harvard.edu/resources/reports_and_working_papers/working_papers/wp1/.
- ⁹ Lipsey, M., Howell, J., Kelly, M., Chapman, G., and Carver, D. (December 2010). *Improving the Effectiveness of Juvenile Justice Programs: A New Perspective on Evidence-Based Practice*. Washington, DC: Georgetown University Center for Juvenile Justice Reform.
- ¹⁰ Jeff Bradach's formulation
- ¹¹ <http://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/>
- ¹² <http://www.whitehouse.gov/blog/2014/07/09/maximizing-impact-social-spending-using-evidence-based-policy-and-low-cost-randomize>
- ¹³ <http://www.whitehouse.gov/blog/2014/07/09/maximizing-impact-social-spending-using-evidence-based-policy-and-low-cost-randomize>
- ¹⁴ <http://coalition4evidence.org/wp-content/uploads/2014/02/Low-cost-RCT-competition-December-2013.pdf>
- ¹⁵ Federal Reserve Bank of San Francisco. (April 2013). "Pay for Success Financing." *Community Development Investment Review* 9(1). <http://evidence-innovation.findyouthinfo.gov/investingEvidence>
- ¹⁶ <http://evidence-innovation.findyouthinfo.gov/investingEvidence>
- ¹⁷ Burwell, S.M., Munoz, C., Holdren, J., and Krueger, A. (July 26, 2013). Memorandum M-13-17: "Next Steps in the Evidence and Innovation Agenda," <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf>.
- ¹⁸ Kania, J., Kramer, M., and Russell, P. (Summer 2014). "Strategic Philanthropy for a Complex World." *Stanford Social Innovation Review*, http://www.ssireview.org/up_for_debate/article/strategic_philanthropy.
- ¹⁹ Viadero, D. (December 2, 2009). "New Head of U.S. Research Agency Aims for Relevance." *Education Week*, <http://www.edweek.org/ew/articles/2009/12/02/13ies.h29.html>.
- ²⁰ Blase, K. and Fixsen, D. (February 2013). "Core Intervention Components: Identifying and Operationalizing What Makes Programs Work." Research Brief. National Implementation Research Network, <http://nirn.fpg.unc.edu/resources/core-intervention-components>.
- ²¹ Child Trends, Inc. (February 2013). "Key Implementation Considerations for Executing Evidence-Based Programs: Project Overview." ASPE research brief. U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, http://aspe.hhs.gov/hsp/13/KeyIssuesforChildrenYouth/KeyImplementation/rb_keyimplement.cfm.
- ²² Emig, C. (February 25, 2011). Personal communication.
- ²³ Grantmakers for Effective Organizations and Council on Foundations. (2009). *Evaluation in Philanthropy: Perspectives from the Field*, <http://www.geofunders.org/resource-library/all/record/a06600000056W4sAAE>.
- ²⁴ McCarthy, P.T. (February 21, 2014). "The Road to Scale Runs Through Public Systems: Smarter Philanthropy for Greater Impact." *Stanford Social Innovation Review*.
- ²⁵ Gopalakrishnan, S. (August 22, 2013). FSG blog post, <http://www.fsg.org/KnowledgeExchange/Blogs/StrategicEvaluation/PostID/478.aspx>

-
- ²⁶ <http://www.carnegiefoundation.org/improvement-research/approach>
- ²⁷ Bryk, A. (March 31, 2011). "It is a Science of Improvement." International Perspectives on Education Reform Group. *Education Week* blog post, http://blogs.edweek.org/edweek/futures_of_reform/2011/03/it_is_a_science_of_improvement.html
- ²⁸ Fulton, K. (2014). Response to "Strategic Philanthropy for a Complex World." *Stanford Social Innovation Review*, http://www.ssireview.org/articles/entry/strategic_philanthropy
- ²⁹ Fulton, K. (2014). Response to "Strategic Philanthropy for a Complex World." *Stanford Social Innovation Review*, http://www.ssireview.org/articles/entry/strategic_philanthropy
- ³⁰ Bryk, A. (March 31, 2011). "It is a Science of Improvement." International Perspectives on Education Reform Group. *Education Week* blog post, http://blogs.edweek.org/edweek/futures_of_reform/2011/03/it_is_a_science_of_improvement.html.
- ³¹ Bryk, A. (March 31, 2011). "It is a Science of Improvement." International Perspectives on Education Reform Group. *Education Week* blog post, http://blogs.edweek.org/edweek/futures_of_reform/2011/03/it_is_a_science_of_improvement.html.
- ³² Benjamin, L.M. and Campbell, D.C. (Spring 2014). "Programs Aren't Everything." *Stanford Social Innovation Review*, http://www.ssireview.org/articles/entry/programs_arent_everything.
- ³³ Preskill, H., Parkhurst, M., Juster, J.S. (2014). "Evaluating Collective Impact: Assessing Your Progress, Effectiveness, and Impact." *FSG Collective Impact Forum*, <http://www.fsg.org/tabid/191/ArticleId/1104/Default.aspx?srpush=true>
- ³⁴ Guijt, I. (2010). "Exploding the Myth of Incompatibility Between Accountability and Learning." In Ubels, J., Acquaye-Baddoo, N.-A., and Fowler, A. *Capacity Development in Practice*, http://betterevaluation.org/resource/overview/accountability_and_learning.
- ³⁵ Lupia, A. (October 17, 2013). "Which Evaluations Should We Believe? Origins of Credibility and Legitimacy in Politicized Environments." Keynote address to the American Evaluation Association plenary session, <http://comm.eval.org/DataVisualizationandReporting/TIGResources1/ViewDocument/?DocumentKey=99c7d088-87be-469b-8d0f-0b0105ae87ed>.
- ³⁶ Kubisch, A. (April 17, 2012). Interview.
- ³⁷ Benjamin, L.M. and Campbell, D.C. (Spring 2014). "Programs Aren't Everything." *Stanford Social Innovation Review*, http://www.ssireview.org/articles/entry/programs_arent_everything
- ³⁸ Smyth, K.F., Goodman, L., and Glenn, C. (October 2006). "The Full-Frame Approach: A New Response to Marginalized Women Left Behind by Specialized Services." *American Journal of Orthopsychiatry* 76(4):489-502. <http://www.childtrends.org/>
- ³⁹ Kania, J., Kramer, M., and Russell, P. (Summer 2014). "Strategic Philanthropy for a Complex World." *Stanford Social Innovation Review*, http://www.ssireview.org/up_for_debate/article/strategic_philanthropy
- ⁴¹ Gibson, C., Smyth, K., Nayowith, G., and Zaff, J. (September 19, 2013). "To Get to the Good, You Gotta Dance With the Wicked." *Stanford Social Innovation Review*, http://www.ssireview.org/blog/entry/to_get_to_the_good_you_gotta_dance_with_the_wicked
- ⁴² Gibson, C., Smyth, K., Nayowith, G., and Zaff, J. (September 19, 2013). "To Get to the Good, You Gotta Dance With the Wicked." *Stanford Social Innovation Review*, http://www.ssireview.org/blog/entry/to_get_to_the_good_you_gotta_dance_with_the_wicked; and Rittel, H. and Webber, M. (1973). "Dilemmas in a General Theory of Planning." *Policy Science* (4), 155-169.
- ⁴⁴ Patrizi, P., Heid Thompson, E., Coffman, J., and Beer, T. (2013) "Eyes Wide Open: Learning as Strategy under Conditions of Complexity and Uncertainty." *The Foundation Review* 5(3).
- ⁴⁵ Kania, J., Kramer, M., and Russell, P. (Summer 2014). "Strategic Philanthropy for a Complex World." *Stanford Social Innovation Review*, http://www.ssireview.org/up_for_debate/article/strategic_philanthropy.
- ⁴⁵ Schall, 1994 and Schon, 1973. Cited in Patrizi, P., Heid Thompson, E., Coffman, J., and Beer, T. (2013) "Eyes Wide Open: Learning as Strategy under Conditions of Complexity and Uncertainty." *The Foundation Review* 5(3).
- ⁴⁶ Walker, D. Response to Kania, J., Kramer, M., and Russell, P. (Summer 2014). "Strategic Philanthropy for a Complex World." *Stanford Social Innovation Review*, http://www.ssireview.org/articles/entry/strategic_philanthropy.
- ⁴⁷ Schon, 1983; Mintzberg, Ahlstrand, and Lampel, 1998; and Mintzberg, 2007. Cited in Patrizi, P., Heid Thompson,

-
- E., Coffman, J., and Beer, T. (2013) "Eyes Wide Open: Learning as Strategy under Conditions of Complexity and Uncertainty." *The Foundation Review* 5(3).
- ⁴⁸Berwick, D.M. (2005). "Broadening the View of Evidence-Based Medicine." *Quality and Safety in Health Care*, 14:315-316 doi:10.1136/qshc.2005.015669..
- ⁴⁹ Bryk, A. (March 31, 2011). "It is a Science of Improvement." International Perspectives on Education Reform Group. *Education Week* blog post, http://blogs.edweek.org/edweek/futures_of_reform/2011/03/it_is_a_science_of_improvement.html
- ⁵⁰ Green, L.W. (March 2006). "Public Health asks of Systems Science: To Advance Our Evidence-Based Practice, Can You Help Us Get More Practice-Based Evidence?" *American Journal of Public Health* 96(3): 406–409.
- ⁵²Swisher, A.K. (June 2010). "Practice-Based Evidence." *Cardiopulmonary Physical Therapy Journal* 21(2): 4.
- ⁵²Thompson, R. and Haskins, R. (Spring 2014). "Early Stress Gets under the Skin." *The Future of Children Policy Brief*, http://futureofchildren.org/futureofchildren/publications/docs/24_01_Policy_Brief.pdf.
- ⁵³ Bryk, A.S. and Schneider, B. (March 2003). "Trust in Schools: A Core Resource for School Reform." *Educational Leadership* 60(6): 40-45.
- ⁵⁴ O'Connell, M.E., Boat, T., and Warner, K.E., eds. (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Report by the National Research Council and Institute of Medicine. Washington, DC: National Academies Press.
- ⁵⁵ Biglan, A., Flay, B., Embry, D., Sandler, I. (2012). "The Critical Role of Nurturing Environments for Promoting Human Well-Being." *American Psychologist* 67(4).
- ⁵⁶ Currie, J. (2013). "Pollution and Infant Health." *Child Development Perspectives*, 7: 237–242.
- ⁵⁷ Rigby, E., Ryan, R.M., and Brooks-Gunn, J. (Fall 2007). "Child Care Quality in Different State Policy Contexts." *Journal of Policy Analysis and Management* 26(4): 887-908.
- ⁵⁸ Green, L.W. and Glasgow, R.E. (March 2006). "Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology." *Evaluation and the Health Professions* 29(1):126-53.
- ⁵⁹ Green, L.W. and Glasgow, R.E. (March 2006). "Evaluating the Relevance, Generalization, and Applicability of Research: Issues in External Validation and Translation Methodology." *Evaluation and the Health Professions* 29(1):126-53.
- ⁶⁰ Furstenberg, F.F. (2011). "The Challenges of Finding Causal Links between Family Educational Practices and Schooling Outcomes." In Duncan, G.J. and Murnane, R.J., eds. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation, pp. 465-482.
- ⁶¹ Gopalakrishnan, S., Preskill, H., Lu, S.J. (2013). "Next Generation Evaluation: Embracing Complexity, Connectivity, and Change." A Learning Brief. FSG, <http://www.fsg.org/tabid/191/ArticleId/964/Default.aspx?srpush=true>
- ⁶² Nelson, D. Interview with Frank Farrow.
- ⁶³ Miller, D. (August 28, 2014). "Don't Let Data Get in the Driver's Seat." Stanford University Graduate School of Business, Center for Social Innovation, <http://csi.gsb.stanford.edu/dont-let-data-get-drivers-seat>
- ⁶⁴ Bryk, A. (March 31, 2011). "It is a Science of Improvement." International Perspectives on Education Reform Group. *Education Week* blog post, http://blogs.edweek.org/edweek/futures_of_reform/2011/03/it_is_a_science_of_improvement.html.
- ⁶⁵ <http://nirn.fpg.unc.edu/learn-implementation/implementation-drivers/integrated-compensatory>
- ⁶⁶ <http://nirn.fpg.unc.edu/learn-implementation/implementation-drivers>
- ⁶⁷ Bryk, A. (March 31, 2011). "It is a Science of Improvement." International Perspectives on Education Reform Group. *Education Week* blog post, http://blogs.edweek.org/edweek/futures_of_reform/2011/03/it_is_a_science_of_improvement.html.
- ⁶⁸ <http://www.carnegiefoundation.org/improvement-research/approach>.