from the **director's desk**

Heather Weiss

As evaluators we are convinced that having the right information available at the right time goes a long way to help organizations improve their program performance and meet their goals. This assumption, however, merits further scrutiny about an organization's readiness to use evaluation information and the evaluation-program relationship. With today's policy and funding environment focused on outcomes, it is crucial more than ever before to not only make sure those outcomes get measured for accountability purposes, but to examine the conditions that facilitate evaluation for continuous improvement and high performance.

In this issue of *The Evaluation Exchange* on continuous improvement in human service organizations, we learn that the management of people in an evaluation can be harder than the management of data. Organizations with a culture that values inquiry, exploration, and discovery, as well as self-examination, are on the forefront of using evaluation to fine tune their programs and be innovative. Where this "culture of inquiry" is lacking, evaluators need to be prepared to adopt new roles as change agents who can dispel fears about evaluation and nurture change in an organization's mindset.

Our contributors assert that for organizations to make use of evaluation findings, the evaluation process needs to be relevant to individuals and to support learning, which leads to new insights and actions within an organization. To this end, evaluators must assume the role of facilitators of learning, engaging participants in a meaningful and non-threatening way to discuss findings and take next steps.

Evaluators are using diverse evaluation approaches and designing tools that are tai-lored to the program or organization being evaluated. Articles in this issue describe various options—reflective assessment, risk analysis, standards of excellence, balanced scorecard, matrices and portfolios, and an outcome-based information system. Evaluators are also striving to clarify their presentations and apply effective training techniques. Finally, they are providing organizations with performance measurement systems to address equity issues through various design options.

A recurrent theme in this issue is that organizations can also go overboard in their focus on outcomes. Our contributors advise, "be realistic and set incremental goals." But even when these are established, evaluators and program practitioners face the challenge of identifying appropriate performance levels and measures for these goals, and determining productive next steps. Understanding the reasons behind evaluation findings about performance is essential to deciding follow-up actions and who should be accountable for what.

It is not enough, though, to think about evaluation for continuous improvement and high performance solely within an organizational context. Evaluators and practitioners need to think about building *a learning system for a field of practice*, such as juvenile justice or full-service schools, and to connect the knowledge gained from individual evaluations to pinpoint gaps, raise new questions, and identify the next generation of best practices.

The HFRP research team and I hope you find this issue provokes new questions and insights. We invite you to share your ideas with us at hfrp@gse.harvard.edu.

Heather

Heather B. Weiss, Ed.D.
Founder & Director
Harvard Family Research Project

> theory & practice

Mindset Matters

Ann Dykman of MPR Associates illustrates that an organization's culture and mindset are important factors in the success of using evaluation for continuous improvement.

Evaluation is essential to any organization's continuous improvement, but eliciting the full cooperation of staff who are unfamiliar with the concept can require more than simply providing top-notch tools and training. Before getting down to the nitty-gritty of performance measures and data collection, evaluators may first have to help create a workplace culture that values inquiry and critical self-review.

In effect, they may have to be cheerleaders for the process itself.

This is a key lesson that evaluation consultants Georgiana Hernández and Mary Visher[1] learned from their involvement in the two-year Working on Workforce Development (WOW) project, sponsored by the James Irvine Foundation in 1998. The San Francisco-based foundation theorized that if nonprofit organizations know how to gather and use performance data, their services will improve. WOW was created to test the theory by sending in expert evaluators to help six Bay Area employment and training organizations refine their performance measurement systems.

Together with Irvine, the evaluators crafted a training plan around several performance measurement workshops and conducted follow-up visits to help agency staff apply the concepts within their own workplaces. Soon after the first workshop, however, Hernández and Visher sensed a challenge to their well-laid plans. Not only were the participating agencies at different levels of readiness for evaluation, they also had varying views about whether the effort was even worthwhile. And, as it happened, motivation had the most impact on an agency's progress.

"It takes a conscious effort to foster agency-wide shifts in mindset, norms, and practices," wrote Hernández and Visher in *Creating a Culture of Inquiry*, the Irvine monograph that details the WOW project. "It takes organizational changes in belief systems about the value of grounding decisions that affect clients in hard data."

In the end, the experiment successfully increased the evaluation capacity of most of the nonprofits, and both the consultants and the foundation felt it yielded valuable lessons for other funders, evaluators, and technical assistance providers.

[1] Hernández is president of Hernández & Associates and Visher is associate director of program evaluation and planning at MPR Associates.

Indicators of Inquiry-Minded Cultures

- Organizations that create new lines of communication across and within their divisions.
- Organizations that move from using performance measurement in one department or division to using it throughout.
- Organizations that revise their performance measurement plan to adapt to new information, conditions, ideas, or policies.
- Organizations that use performance findings as a springboard for discussions about their values, mission, and hidden assumptions in their practices.

Source: Hernández, G., & Visher, M. G. (2001). *Creating a culture of inquiry*. San Francisco: The James Irvine Foundation. Available at www.irvine.org.

Lesson One – Take Time to Build Relationships and Trust

Hernández and Visher visited each agency twice in the year before the training was set to begin. They interviewed managers, collected data where possible, toured the facilities, and talked to other staff to aid their understanding of the agencies' capacity for self-evaluation. These initial assessment visits gave the evaluators time to build rapport with participants and provided a baseline against which to measure progress. It was during this stage that Hernández and Visher first realized their training plan had to change.

"We *thought* we could go in cold," without much of an understanding of each agency's culture, Visher recalls. "We learned that it takes time for an organization's employees to trust you." For example, agency employees were concerned about opening their performance up to scrutiny in case that might influence their relationship with the foundation. "Early on we had to ensure them that the outcome data reported in their performance measurement systems would not be used as a basis for funding," Visher says.

In retrospect, both the consultants and the foundation also questioned whether offering grants was a good idea, given that the most important determinant of commitment and progress turned out to be an agency's enthusiasm for the project rather than grant size, organization size, stability, finances, or reputation.

Lesson Two – Involve More Staff Throughout the Process

The foundation and the consultants had already decided on the structure for the technical assistance before they first visited the project participants. Soliciting more upfront involvement from the participants would have moved the project along much faster, say Hernández and Visher. Instead it took many months to get everyone on board.

The WOW organizations that were most successful in developing their performance measurement plans seemed to be those that rallied the broadest participation of staff, while keeping midlevel and top managers active throughout the process.

Lesson Three – Clarify Roles, Responsibilities, and Expectations

"We should have laid our cards out from the beginning and scaled back on the more grandiose, overreaching goals [such as expecting to see improved performance at the end of two years]," says Martha Campbell, Irvine's director of evaluation. "Instead, what you need are shorter-term, mutual expectations."

Both the foundation and the consultants suggest that these steps be followed in any performance measurement project:

- Ask each organization to set up a project team, with a lead person, to spearhead the planning process.
- Develop a joint memo of understanding that articulates the project purpose, the expected "deliverables," the project timeline and time requirements, and the roles and responsibilities of designated agency staff and technical assistance providers.
- Make sure everyone agrees on the level of staff time to be invested.

Hernández stresses the need for "constant reality checks" that restrain overly enthusiastic agency staff who want to take on too much. For example, three directors decided to develop plans that encompassed not only employment and training services, but their entire agency operations—including facilities and operations, administration and fundraising, and for-profit business divisions for those who had them. Others wanted to document and measure every single action case managers took every day.

"There can be an excitement about what can be measured, but it can be exhausting to do it all," Hernández says. "In our follow-up conversations we learned that some of the agencies have dropped half of the indicators they came up with."

Lesson Four – Create a Schedule That Works

Because of potential staff turnover and the differing levels of commitment among agencies, the evaluation consultants found they got better results when they worked intensively with participants over shorter periods instead of visiting them less frequently throughout the year. But this is not a rigid rule. Ultimately, the organization that is trying to measure and improve its performance should determine the pace that it feels comfortable with, and then develop a schedule and commit to it.

Lesson Five – Nurture Deep Cultural Changes

To their surprise, Visher and Hernández found they had to spend a good deal of time nurturing change in the organizational mindset around the value of self-evaluation as they were helping organizations improve their technical capacity to use data. They had to get agency staff interested in thinking about how to quantify their day-to-day work and how to ground programmatic decisions in hard data. This was a new approach for many and for some it felt quite threatening.

Visher suggests an initial brainstorming session that teases out the types of work staff members do each day with clients. These are the practices that will be linked with desired outcomes. Once staff begin to see their work being valued, Visher says the change in mindset is so tangible "you can almost hear the click." As the evaluation consultant continued to work with the agencies on their performance measurement system, they began to identify the changes in mindset and to foster them (see box on page 2).

Continuous improvement depends on the adoption of performance measurement as a core value, says Hernández. "I believe wholeheartedly that the process helps staff chart and then move along the same path within the organization, so in the end, managers don't have to spend so much time handling crises and putting out fires."

Ann Dykman
Senior Associate for Communications
MPR Associates, Inc.
77 Warren St. #5
New York, NY 10007
Tel: 212-608-5567
Fax: 212-608-2557
Email: adykman@mprinc.com

> ask the expert

What Is a Learning Approach to Evaluation?

Rosalie T. Torres, Ph.D. is Director of Research, Evaluation, and Organizational Learning at the Developmental Studies Center in Oakland, California. Her 24-year career in evaluation has focused on researching, teaching, writing about, and practicing a learning approach to evaluation.

A learning approach to evaluation is contextually-sensitive and ongoing, and supports dialogue, reflection, and decision making based on evaluation findings (Torres & Preskill, 2001). Its primary purpose is to support learning that can ultimately lead to effective decision making and improvement in department, programmatic, and organization-wide practices. It is based on the premise that to achieve learning and improvement, evaluation must:

- Consider (but not be completely governed by) the context in which it is being conducted (i.e., stakeholders' information needs, political realities, organizational culture).
- Be conducted routinely.
- Provide opportunities for stakeholder participation—in particular, opportunities to ask questions about, discuss, and reflect on the meaning of evaluation findings.

A learning approach can be taken with any kind of evaluation. The goal is to conduct the most valid, relevant, and credible evaluation possible—given the information needs and the existing political, logistical, and resource constraints—*and* to actively facilitate *learning from* the evaluation and taking action. This means establishing a balance between accountability and learning roles for evaluation.

Learning occurs best among individuals who regard the information they are reviewing (i.e., evaluation findings) as credible and relevant to their needs. Involving stakeholders in designing and conducting an evaluation helps assure their ownership of, and interest in, its findings.[1] Learning also occurs best among individuals who have an opportunity to ask questions about evaluation methods, consider other sources of information about the topic in question (including their own direct experiences), and at the same time hear others' perspectives.

One way to take a learning approach with any evaluation, whether you have been involved in conducting it or not, is to enable (or plan for) deep consideration of its findings and the planning of next steps. Often evaluators and their clients think the major work of an evaluation is complete when the final evaluation report has been written. To the contrary, some of the most productive work of an evaluation occurs when stakeholders convene to review and discuss findings and implications. The meet-

[1] See Patton, 1997 and Preskill & Torres, 1999 for more about stakeholder involvement in all phases of evaluation.

ing should be conducted with the key principles of adult learning and staff development in mind: have participants engage with the learning material; see, hear, and do something with its content; and integrate new knowledge with what they already know (Torres, Preskill, & Piontek, 1996).

Consider these guidelines for working sessions of several hours or even over the course of a daylong retreat to review, discuss, and interpret evaluation findings:

1. Begin with a presentation of findings via posters, overheads, a PowerPoint presentation, or other form of visual display so that participants receive and process the findings together rather than individually.
2. Give sufficient opportunity for participants to ask questions about the evaluation procedures and data analysis.
3. Provide an interactive experience for participants to engage with the findings. For example, have participants work in small groups to develop conclusions based on the findings. (The groups might each take a different section of the findings.)
4. Have each group present their conclusions, and give the evaluator an opportunity to comment on or raise questions about conclusions that may have been missed or overstated.
5. Post findings and conclusions around the room and have groups of participants visit each poster, recording their thoughts about implications and actions to consider.
6. Debrief in a large group discussion on the implications and actions.
7. Conclude with a more detailed action-planning activity or definite next steps which will lead to action planning.

Rosalie T. Torres, Ph.D.
Director of Research, Evaluation, and Organizational Learning
Developmental Studies Center
2000 Embarcadero, Suite 305
Oakland, CA 94606
Tel: 510-533-0213 x235
Email: rtorres@devstu.org

References

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.

Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation, 22*(3), 387-395.

Torres, R. T., Preskill, H., & Piontek, M. E. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. Thousand Oaks, CA: Sage.

Increase Equity Without Losing Efficiency

Professor of Public Administration, Joseph Wholey, explains that contrary to popular thought, it is possible to increase program equity without compromising program efficiency—through performance measurement and management systems.

Equity—the fairness with which organizations allocate resources, deliver services, or achieve outcomes—is an important value for both public and nonprofit organizations. It is often suggested that equity and efficiency are incompatible, and that efforts to improve program effectiveness run counter to efforts to increase equity. For example, a job training program may enroll only the most job-ready clients in an effort to increase its number of placements. This approach will increase the program's efficiency, but reduce equity in the delivery of its services. In situations like this, being *results-oriented* conflicts with being *equitable.*

Program equity and efficiency do not have to be incompatible. If policymakers and senior managers agree that equity is important, they can: first, define equity goals, second, develop performance measurement systems that emphasize both equity and efficiency, and third, adjust management systems to create greater equity in their programs and in the communities that they serve.

Defining Equity Goals

The first step to increasing equity is achieving agreement among stakeholders on what the equity goals are. Evaluability assessment, a process for determining whether a program is ready for useful evaluation work, can help create agreement on equity goals. Through this process, policymakers and program managers can: compare and contrast the performance expectations of key stakeholders; reach a reasonable level of agreement in defining the program's intended inputs, activities, outputs, and outcomes; establish or revise the program's goals; and develop or revise the strategies for achieving these goals.[1]

Involving stakeholders who represent different population groups in this process helps organizations to develop goals and strategies that include a clear focus on equity. If key stakeholders emphasize equity in service delivery or service outcomes, then the program's goals should reflect that emphasis.

Measuring Progress in Creating Equity

Organizations have many options regarding what types of performance measurement systems they can use to address equity concerns.

Organizations can create performance measurement systems that directly measure equity. For example, an organization may use role-playing to evaluate whether different groups of job applicants, home buyers, or renters receive equal treatment. "Paired role-playing has been used since the 1970s by private and public fair housing organizations to test for discrimination against minority home seekers by real estate and rental agents."[2]

Organizations can use performance measurement systems that report separate performance scores for different population groups. For example, to encourage effective services to disadvantaged populations, many school districts now measure and report both overall student performance and the performance of various minority groups. Such efforts will grow as the federal government encourages and requires state and local educational efforts that "leave no child behind."[3]

Organizations can use statistically-adjusted measures of effectiveness or statistically-adjusted performance standards that reflect the influence of client characteristics. For many years, for example, the U. S. Department of Labor has adjusted performance standards for job training programs to reflect the influence of client characteristics and other degree-of-difficulty factors.[4]

Adjusting Management Systems for Equity

Once organizations have set equity goals and developed performance measurement systems that address equity concerns, they can develop results-oriented management systems to help increase equity. Though attention often tends to focus on equity of inputs, the most interesting and challenging issues typically relate to equity in service delivery or outcomes.

Individuals and organizations seeking to increase equity must find ways to change management systems to create the incentives needed to change behavior. If the incentive systems don't change, nothing will change. Intangible incentives may include regular

[1] See Related Resources: Wholey, J. S. (1983); Wholey, J. S. (1994).

[2] Turner, M. A., & Zimmerman, W. (1994). Acting for the sake of research: The use of role-playing in evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 310-337). San Francisco: Jossey-Bass.

[3] Under the No Child Left Behind Act of 2001, each state will: measure the progress of all students toward state-developed standards for what a child should know; the progress of each state, school district, and school will be publicly reported; and state and district reports will show achievement gaps for students who are economically disadvantaged, are from racial or ethnic minority groups, have disabilities, or have limited proficiency in English (U. S. Department of Education. (n.d.) *Introduction: No child left behind.* Retrieved July 18, 2002, from www.nochildleftbehind.gov/next/overview/index.html).

[4] U. S. General Accounting Office. (1999). *Managing for results: Measuring program results that are under limited federal control.* Washington, DC: Author.

Related Resources

Wholey, J. S. (1983). *Evaluation and effective public management.* Boston: Little, Brown.

Wholey, J. S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15-39). San Francisco: Jossey-Bass.

Moving Ahead on Results Accountability:
The Human Element

Sara Watson of the Pew Charitable Trusts explains that a results accountability system must extend beyond the purely technical to also address the management of people.

When *The Evaluation Exchange* published its first newsletters on results accountability in the mid-1990s, federal, state, and local governments were just beginning to introduce these concepts into child and family services. Since then, researchers and practitioners in results accountability have produced a huge amount of information on how to define, measure, and collect the data necessary for accountability systems.

Hundreds of state and local government reports and websites provide extensive guidance on choosing and measuring results, including lists of results and benchmarks, language frameworks, and definitions for all of the varied categories of measures. The various taskforces that are charged with creating a results accountability system have spent years developing their own answers to the questions of what results to seek, what nomenclature to use, what numbers to collect, and how they should be collected. This level of sophistication and variety is important—

without the right foundation, results accountability systems will be at the best ineffective and, at the worst, destructive.

Answering the Difficult Questions

But these are only the initial questions that must be addressed in order to use results to actually improve the well-being of children and families over time. *Much* harder are the questions about people, not numbers—about management, not measurement. Once the numbers are in hand, what should be done with them? Who should be accountable for what? How does one unearth the reasons behind a particular performance? What's the best way to decide what follow-up actions will help the numbers improve? What process is both fair to all parties and effective at promoting continuous improvement? What actions will engage workers as partners and encourage them to do their best, instead of making them feel criticized and threatened?

These questions need to be asked, and answered, at the level of individual managers and workers, as well as for federal and state officials and Congressional and state legislative committees. The field needs to apply its considerable intellectual capacity to these questions if results accountability is to become embedded in the fabric of our formal and informal supports for children and families.

An Example of Addressing the Human Element

Missouri's Caring Communities initiative is one of the earliest and best-known community-based results accountability initiatives in the country. Last year the Family and Community Trust (the organization that leads Caring Communities)[1] began a process to design a true shared accountability system, with performance targets and consequences.

One of the most important parts of this system was a formal set of responsibilities for all three of the main partners: the state agencies, the community partnerships, and the Family and Community Trust. While there was a great deal of precedence for its list of performance targets, there was much less written in the field to guide the development of shared responsibilities. One of the most important breakthroughs came when the agency staff said "we are accountable too" and backed that up with specific tasks that they would carry out to support the partnerships' performance on results. Addressing those very human elements, at the same level or more than data concerns, was key to moving ahead with the de-

[1] As this issue was going to press, the state was disbanding FACT. However, the experience was still important in developing an accountability design accepted by all the partners.

Examples of Ways to Support Improvement

Private reward and pressure – private conversations between those responsible for performing and those who hold them responsible

Public reward and pressure – publicizing performance data and allowing public praise or inviting public inquiry

Tangible rewards for success – money, equipment, etc.

Increased autonomy – more freedom from certain rules or requirements (e.g., administrative paperwork)

Increased assistance – voluntary or mandatory assistance following poor performance

Reduced autonomy/increased oversight – decreased freedom and new or stronger oversight procedures

Reduced or transferred funding – less or no financial support for poor performance (a late-stage option that must be administered in ways that protect vulnerable children and families)

Changing personnel – moving people who were responsible for improvement, but not able to produce it (a last resort option and again administered with care)

Source: Watson, S. (2000). *Informed consent: Advice for state and local leaders on implementing results-based decisionmaking.* Washington, DC: The Finance Project. Available at www.financeproject.org/informed_consent.htm.

Helping Nonprofits Strive for High Performance

Amy Coates Madsen describes how, by setting best practices for nonprofits, the Standards for Excellence program both helps nonprofits to improve and increases public confidence in them.

Aprogram director has just been promoted to executive director and wants to learn about how her organization's management and governance practices measure up to best practices within the sector. In the past, there was no easy way for her to benchmark how her organization's practices compared to norms in the nonprofit sector, let alone to know whether such norms existed.

That changed in 1998 when the Maryland Association of Nonprofit Organizations (Maryland Nonprofits) launched the Standards for Excellence program, a comprehensive ethics program that works to strengthen a nonprofit organization's management and leadership, while increasing the public's confidence in the nonprofit sector. The program is now being replicated in nonprofit associations in Pennsylvania, North Carolina, Georgia, Louisiana, and Ohio.

The program includes four components: 1) the Standards for Excellence, 2) nonprofit education about the standards, 3) voluntary certification of organizations meeting the standards, and 4) public education about nonprofit ethics and accountability.

1. The Standards

The Standards for Excellence is a code of conduct that covers eight major topic areas, including:

- Mission and program evaluation
- Governing body
- Conflicts of interest
- Human resources
- Legal and financial issues
- Openness and disclosure
- Fundraising
- Public policy/public affairs

Best practices in nonprofit organizations are outlined and carefully laid out in 55 performance indicators or standards.

Maryland Nonprofits encourages nonprofits to use the Standards for Excellence as a self-assessment checklist for ethical nonprofit operations. Most nonprofits implement about 45 of the 55 standards. Once a nonprofit organization conducts their baseline organizational assessment, the organization is encouraged to devise a plan to implement their remaining unmet standards. Implementing them is a comprehensive process involving both the organization's board and staff.

2. Nonprofit Education

Maryland Nonprofits offers a wide variety of resources to help nonprofits implement the Standards for Excellence. Written educational materials, which include instructional pieces on each best practice, sample policies, procedures, and excerpts from some of the best experts in the fields, have been developed for each area.

Special courses also help introduce the practices and ideals espoused by the Standards of Excellence. One of the association's most popular programs is a four-session clinic series where nonprofit leaders come together to learn about the details of the standards and engage in peer-to-peer exchange, a process to help organizations garner suggestions from their colleagues. In Maryland alone over 400 organizations have taken advantage of training courses to improve their operations.

3. Voluntary Certification

For nonprofits interested in adhering to all 55 performance indicators, Maryland Nonprofits offers a voluntary certification program. To date, 25 organizations have earned the Seal of Excellence for successfully completing this staff-assisted, peer review process.

4. Public Education

The Standards of Excellence public education effort informs the public about characteristics of well-managed, responsibly governed nonprofits. The association works to educate the public through ongoing media outreach, a public service announcement campaign, and by providing packages of easily adaptable promotional material to certified organizations for use with their donors and the public.

The Results

Do these efforts make a difference in strengthening nonprofit performance? A survey of over 350 nonprofit organizations shows that those with access to the services of the Standards for Excellence program have strengthened their governance, management, and organizational practices to a greater degree than organizations without access to the program. In addition, these organizations not only have higher expectations for their performance than the other organizations, but they also outperform them.

Amy Coates Madsen
Program Director
Standards for Excellence
Maryland Association of Nonprofit Organizations
190 West Ostend Street, Suite 201
Baltimore, MD 21230
Tel: 410-727-6367
Fax: 410-727-1914
Email: acmadsen@mdnonprofit.org

For more information about the Standards for Excellence program visit www.standardsforexcellence.org. For more information about Maryland Nonprofits visit www.mdnonprofit.org or call 410-727-6367.

Reflective Assessments: A Tool for Learning

Sharon Edwards and Ira Cutler of Cornerstone Consulting Group explain how organizations can use reflective assessments to assess their progress and consider the choices ahead.

In times like these of diminished resources, most funders expect the organizations they fund to evaluate their programs in some fashion. In some instances, particularly when organizations hope to establish a standard of practice or want to replicate their program, a formal, structured third-party evaluation, with appropriate controls, random selection, careful data collection, and measurement over time is the best method.

Many organizations, however, do not need and cannot afford an extensive third-party evaluation. Instead, their need is for a careful process that will provide them with the opportunity to stop and carefully reflect before moving on to the next stage. The best tool to use in this case is a reflective assessment.

Reflective assessments entail a short and inexpensive process through which organizations can assess their progress, indicate the lessons they have learned, document their accomplishments, and consider the choices ahead. Reflective assessments have been used in a number of ways by a variety of organizations:

- The Ford Foundation's Community Youth Development Initiative and the Pew Charitable Trusts' Neighborhood Preservation Initiative used reflective assessment to produce lessons learned reports, documenting the accomplishments of their initiatives for a broad audience.
- The Annie E. Casey Foundation's Rebuilding Communities Initiative used reflective assessment to give a voice to participants' view of the initiative to supplement the formal reports of external evaluators.
- The Los Angeles Children's Policy Council (LACPC), approaching a 10-year milestone, used a reflective assessment process to build consensus and frame discussions about plans for the next 10 years.

Each reflective assessment is a separately designed project, but generally unfolds in five distinct phases. These phases are outlined below, using Cornerstone Consulting Group's reflective assessment of LACPC as an illustrative example.

Related Resources

The Cornerstone Consulting Group. (2002). *End games: The challenge of sustainability*. Houston, TX: Author. Available at www.cornerstone.to/what/publications.html.

The Cornerstone Consulting Group. (2001). *Communities and youth development: Coming together*. Houston, TX: Author. Available at www.cornerstone.to/what/publications.html.

Phase One – Define the Engagement

In the first and in some ways most important phase, Cornerstone consultants meet with the leadership group to define the engagement: What is the purpose of the assessment? How will the findings of the assessment be used? Who is the audience? The LACPC sought a "lessons learned" report to inform the field, but, after a series of frank discussions, it decided on a different, more internally focused approach. Rather than focusing on exporting ideas, the LACPC decided it needed to take a deep organizational breath and think hard about options and goals for its next 10 years. Had the consultants not taken sufficient time at the front end to define the engagement, they might have produced a report to serve one purpose, but they would have missed the most important mark.

Phase Two – Establish an Information Base

Using readily available information, including annual and progress reports, mission statements, operating manuals, and formal evaluations, Cornerstone establishes an information base. In the case of the LACPC there was no lack of written information and, in fact, the challenge was figuring out what mattered most.

Phase Three – Capture Individual Perspectives

The heart of the reflective assessment process is a series of structured interviews with project leaders, stakeholders, and other constituency groups intended to capture individual perspectives. Cornerstone consultants spoke with 20–25 key informants in Los Angeles County for the LACPC reflective assessment. These interviews offer anonymity and time for personal reflection. Through dialogues with those closest to the effort, Cornerstone develops an analysis of achievements, challenges, the elements stakeholders consider most successful, and areas they see as problematic. The diversity of perspectives provides a voice for all groups involved in the effort. The interviews initiate and set the tone for a process of discovery and exploration and interviewees often say that the interview itself is an enjoyable learning experience.

Phase Four – Prepare an Assessment Report

Cornerstone prepares an assessment report providing an informed view of the initiative's status and direction, and what has been learned during its course. This document is presented for review as a check on accuracy and tone before being finalized for broader distribution. Consultants ask if their picture is accurate, seek to clarify differences, and try to establish common ground. The report developed for the LACPC, *The Tasks Ahead*, raised questions that resonated with many CPC (Children's Policy Council) members and created a "buzz" that focused attention on important issues.

Phase Five – Present the Report

On the basis of feedback and discussions, Cornerstone finalizes the report and often makes a presentation in a formal setting,

Risk

John Bare of the Knight Foundation shares his foundation's definition of the term "risk" when it comes to investing in initiatives, borrowing from the language of money managers.

I'm a little nutty when it comes to risk. I go looking for fried chicken, which increases my chances of heart disease, but alter my path to avoid walking over city sidewalk grates. I'm scared of being the million-to-one rube pulled from the sewer.

At least I've thought it through. In philanthropy, we frequently use the term "risk" without explaining what we mean.

At the John S. and James L. Knight Foundation in Miami, where I work as director of program development and evaluation, the strategic plan that trustees adopted in 2000 says the 26 U.S. communities where we make grants "will be considered laboratories, where high-risk and experimental ventures are tested and approaches that have been demonstrated to work are implemented and nurtured."

Our staff loved the language, but it put us on the spot. We needed a common understanding of "risk." We got help from evaluator Michael Quinn Patton who encouraged us to borrow concepts from money managers. Some grants are sort of like long-term bonds, some are blue chip stocks, and so on. From that, we developed ways to describe risk and manage it across a portfolio of grants.

Up front, we work with nonprofit partners to agree on a grant's anticipated outcome. In financial terms, this is the ROI—return on investment. In recent years we added this explanation to the grant summaries we write for our board. This sharpens the grant development arguments. Are we assuming risk to produce effects that benefit individuals directly? Or to help communities learn something? Or both?

Caroming through trial-and-error innovation may not be the surest way to provide immediate benefits to individuals, but the promise of knowledge that benefits communities is a counterweight.

"We've learned that invention requires failing, starting over, learning from the mistakes, trying again," our annual report
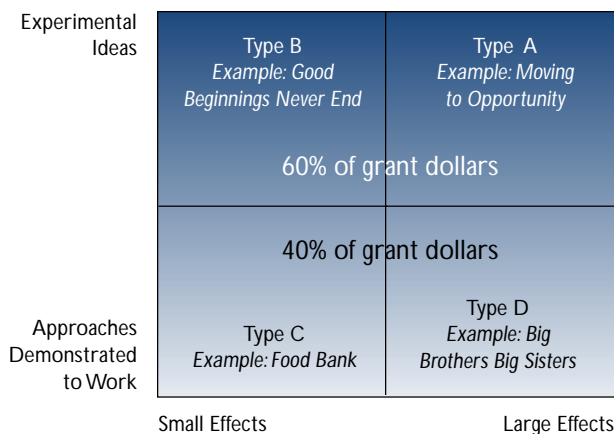
Hypothetical Risk Tolerances

The chart depicts tolerances for "idea risk" by a hypothetical foundation interested in investing 40% of its resources in relatively safe approaches and 60% in experimental ideas. Grant investments fit along three dimensions: 1) the magnitude of the intended benefits (horizontal axis), 2) the degree of risk associated with the idea (vertical axis), and 3) the potential for gaining new insight into effective practice (light-to-dark shading).

Experimental Ideas   Evaluation is a priority. Assessing process and program outcomes is critical to learning what worked, what didn't, and why.

Type A: Relatively high risks are offset by the potential for large benefits. Requires long-term investments. *Example: Moving to Opportunity – a national experiment using housing vouchers to help poor families secure homes in safer neighborhoods.*

Type B: Investments dip a toe into uncharted waters. The investment and the expectations for effects are small. Incremental gains may build toward something larger. *Example: Good Beginnings Never End – a door-to-door training service in Long Beach, California that lends a hand to adults who run child-care programs in their homes.*

Experimental Ideas	**Type B** *Example: Good Beginnings Never End*	**Type A** *Example: Moving to Opportunity*
	60% of grant dollars	
	40% of grant dollars	
Approaches Demonstrated to Work	**Type C** *Example: Food Bank*	**Type D** *Example: Big Brothers Big Sisters*
	Small Effects	Large Effects

Approaches Demonstrated to Work   Evaluation of these programs monitor implementation and the community conditions in which practitioners are working.

Type C: Short-term investments where the benefits are small, but relatively certain. Are critical for sustaining momentum in communities. Benefits are often highly visible. *Examples: Food banks and coats-for-kids programs.*

Type D: Have strong track record of producing relatively large effects. Require long-term investments. *Example: Big Brothers Big Sisters program.*

A conversation with
Paul Light

Paul Light is a Senior Fellow of Governance Studies, Director of the Center for Public Service, and a Senior Adviser to the Presidential Appointee Initiative at the Brookings Institution in Washington, D.C. He is also an instructor at the John F. Kennedy School of Government at Harvard University. Previously he was Director of the Public Policy Program at the Pew Charitable Trusts. He has written 14 books, including most recently _Pathways to Nonprofit Excellence_.[1] Light received his Ph.D. from the University of Michigan.

Q **Your new book, _Pathways to Nonprofit Excellence,_ asserts that continuous improvement is at the core of what it means to be a high-performing organization. How can evaluation help nonprofits to be high performers?**

A There are three questions that all organizations must address, whether they're nonprofit, private, or governmental:

1. Why do they exist?
2. Whom do they serve?
3. How will they know they are successful?

I believe that rigorous measurement should be at the center of a nonprofit's mission. If you don't have some sort of measurement system for tracking activities, outputs, and hopefully outcomes, you're not going to know if you are on the right track. Every nonprofit that I've come to admire over the years has started with a commitment to rigorous measurement as part of their core ethic. When I ask how they will know if they are successful, they are able to articulate clearly the outcomes and measures they use to assure they are headed down the right path.

Also, a commitment to careful measurement and evaluation is essential to the task of innovation. Most innovation does not involve "mad-scientist thinking." Innovating organizations correct themselves over time and fine tune their programs through careful evaluation. They have an extraordinary commitment to measurement because they know that: a) if they are going to challenge prevailing wisdom through innovation, they better be prepared for the prevailing wisdom to push back and challenge them to prove that their new ideas work, and b) their new ideas are not going to work unless they commit themselves to trial and error, which is guided by careful evaluation.

Q **How is the "outcomes movement" affecting the performance of nonprofits?**

A I think the outcomes movement, with its focus on measuring real outcomes rather than activities or short-term outputs, is very favorable, as long as it is not focused only on "compliance." I think the best outcomes measurement is

> _A commitment to careful measurement and evaluation is essential to innovation. Most innovation does not involve "mad-scientist thinking." Innovating organizations correct themselves over time and fine tune their programs through careful evaluation._

developed by the nonprofit sector in collaboration with the party requesting it, rather than imposed from the outside.

Also, outcomes measurement should not be another paperwork exercise that doesn't matter in any tangible way to the allocation of dollars and resources. For the last 10 years outcomes have been pushed very hard at the federal level with, I think, mixed results because in some cases it has not really mattered for the things that federal agencies pay attention to—budget, personnel, financial rewards for personnel, etc.

We interviewed federal employees over the last six months about morale. One federal employee talked at length about how his agency has done such a good job at producing strategic plans and developing very detailed metrics for assessing impacts, and that all that has occurred is that Congress has more ammunition with which to pummel away at his agency. The reality is that you can have very high performance and not get a budget increase, and you can have very poor performance and get a significant budget increase.

The outcomes movement is going to have the same results in the nonprofit sector if it is not intimately tied to things that matter to nonprofits—meaning dollars for programmatic activity—and if it is not done fairly.

Q **Can you give examples of federal agencies that have used outcomes in a meaningful way?**

A The Department of Transportation (DOT) is considered the best-run federal department in terms of outcome measurement. The Federal Highway Administration (FHA) within DOT in particular has set very broad strategic goals for itself that deal with things like reducing highway fatalities. That's a very difficult outcome to control, but the FHA has embraced it and said, "If we are going to do our job, we're going to reduce highway fatalities, and we are going to do that by increasing the use of seatbelts, getting better at enforcing speed limits, etc."

[1] Light, P. C. (2002). _Pathways to nonprofit excellence._ Washington, DC: Brookings Institution Press.

DOT is setting a very high mark in saying, "You've got to take responsibility for some outcomes that may be just beyond your reach." Traditionally, the agency would say the number of interstate highway deaths is not something they could control. But DOT felt their rules on seatbelt use, crash survivability, airbags, etc., were designed to make accidents more survivable and so they should hold themselves accountable for that. This approach can really energize an agency and a workforce. Every day that FHA staff go into work they should be thinking, "We are about saving lives in this agency."

Q Many nonprofits feel they do not have the time, staff, and other resources for evaluation. What would you say in response?

A I understand the pressures that nonprofits face, especially small ones for whom the choice to evaluate may mean cutting back on desperately needed staff or technology. But I think evaluation is a wise investment. I'd say to small, struggling nonprofits, "If you're coming to work every day because of the mission, you need to have some understanding of whether you're actually making headway. And that's got to be more than hunch and intuition."

Funders can help by not cutting their evaluation and capacity-building budgets in times of retrenchment. Capacity building and evaluation are usually the first things to go. Foundations say what their grantees say, "If we have to make a choice between evaluation or capacity building and saving trees, for example, we're going to focus on saving trees." I think that could be, even though it's not a proven fact, a terrible investment decision.

It denies the possibility that investments in evaluation and capacity more generally can improve productivity and increase efficiencies in ways that will stretch those dollars much further than they would otherwise go. It could be that during periods of retrenchment, the most important investment foundations can make is in organizational capacity building, precisely because that's what many organizations cut first.

There was a survey of California nonprofits last fall in the wake of September 11. Most nonprofits contacted were already starting to cut back on staffing, MIS investments, pay raises, bonuses, and new benefits programs. I spent some time in California over the last two months and you can see the effects of those cutbacks. Organizations became more stressful places to work, started to burn out their employees, and saw higher turnover. That can't be good for saving trees in any respect.

Q How should capacity-building efforts be evaluated?

A We are working very hard on that question right now. The traditional view has been to evaluate capacity building in terms of its outputs. It's relatively easy to look at capacity building on fundraising, for example, and mea-

Paul Light

sure the number of letters that went out, the number of successful contacts, and the amount of money brought in. Periods of retrenchment that focus on outputs alone almost encourage funders to reduce their capacity-building support because it's something that they believe can be delayed.

The real challenge in evaluating capacity building is to demonstrate its impact on actual programmatic outcomes. It's very difficult to do. The evaluation community has not been very effective at designing evaluations that ask how capacity building improves the efficiency or productivity of an organization in actually doing things like saving trees, educating children, feeding the hungry, supporting HIV/AIDS patients, and so forth.

We need to know, "What are the linking mechanisms that come out of capacity-building efforts that produce more trees saved?" For example, in theory, an MIS should increase productivity, and productivity is a linking mechanism that should increase programmatic impact. Efficiency, strategic alliances, morale, and job satisfaction are other linking mechanisms. Even though this makes intuitive sense, we haven't done a good job of proving it.

Q What is currently missing in evaluations of the nonprofit sector? What might be future directions?

A I think evaluators generally neglect organizational variables as possible causes of programmatic failure. Program evaluators tend to look at program design as the most important factor in whether a program succeeds or fails. Evaluations should also look at staff training, organizational configuration, access to technology, and everything from supplies to strategic planning.

We've got a bunch of evaluators who do *program evaluation*; we've got a very small number of evaluators who do *capacity-building evaluation*. I'd like to see more interaction between the two, because I suspect their interaction could result in some methodology to figure out how to attribute program failure and success to program design and organizational capacity.

Also, a barrier to evaluation is that many organizations think it's a very costly enterprise that's going to take a long time. And we all know better. But perhaps the evaluation community can come up with a different, deeper set of products that can help organizations at different points in their lifecycle, and help organizations with varying amounts of money. This is not to be a substitute for full-blown, empirically anchored evaluation, but is to get organizations to think about measurement more effectively. Evaluators need to think about different levels of evaluation and ways to get organizations to think of evaluation as something that's a constant benefit rather than an eventual trial.

Julia Coffman, Consultant, HFRP
M. Elena Lopez, Senior Consultant, HFRP

Participatory Evaluation for Continuous Improvement

Kim Sabo of Sabo Consulting and Dana Fusco from York College, CUNY illustrate how they conducted a participatory evaluation of an after school literacy initiative to support its continuous improvement.

In 1998, we were commissioned to evaluate a collaborative of 10 community-based after school programs in New York City that had just instituted a literacy initiative. Since the programs were small and were interested in continuous improvement, we thought a participatory evaluation (PE) would be the best method to use.

Working with the collaborative's literacy coordinator, the initial step was to create a logic model of the entire literacy initiative. We quickly learned that there were many different understandings and practices of literacy across the 10 after school programs. If PE was going to offer programs the opportunity for continuous improvement, then we had to create strategies to evaluate the initiative while staying true to the uniqueness of each program within the initiative.

One strategy was a Literacy Program Model (LPM). The LPM allowed each program to identify three of their "best" literacy practices and then position them in relation to the program's objective(s), resources, and assessment tools. In developing the LPM, staff began to recognize the areas within the program that could be strengthened, i.e., which objectives were not being met. Staff defined a Plan of Action that specified next steps with regard to adding or changing literacy activities, acquiring additional resources, and developing documentation and assessment

strategies. In this way programs used the LPM to guide continuous improvement of program planning.

The extent to which the Plan of Action was implemented became a central focus of both individual program evaluations and the evaluation of the overall initiative. Next we needed methods to document the progress of the programs and the initiative.

Two User-Friendly Data Collection Tools

Throughout the PE process, two data collection tools were especially helpful: the House Matrix and the House Portfolio.

The House Matrix is a program overview organized according to seven areas of potential change: philosophy/mission, staff development, facilities, literacy practice, assessment, parent involvement, and school/community collaborations. Each matrix includes specific program information (mission statement, number of staff workshops, child-staff ratio, academic and enrichment activities, assessment procedures, number of parent meetings and use of parent advisories, and school and community partnerships). Data for the house matrices are gathered, reviewed, and revised in collaboration with staff. This organizational tool provides a contextualized snapshot of each program that helps both evaluators and staff members understand their program better.

The House Portfolio expands on the information represented in the House Matrix, with the seven areas of change sectioned off in a three-ring binder. All staff collect data for the portfolio and periodically come together to reflect on the findings. The portfolios help individual programs, the literacy coordinator, and the evaluation team understand their progress and accomplishments over the course of a year. In addition, the portfolios become a communication tool, allowing staff to share the program's overall work and development with funders, parents, students, and other stakeholders.

The Results

By the end of the evaluation, all programs were able to collect data and were beginning to understand the usefulness of evaluation and reflection. Even when programs valued the long-term gains of PE, high turnover rates often made the directors wary of investing in further staff development. Thus, it is unclear whether we could have fully built evaluation capacity within each of the programs. What we were able to do was address each of our evaluation questions while simultaneously supporting programs to gain a better understanding of the relationship between program planning and evaluation.

Kim Sabo, Ph.D.
Consultant
Kim Sabo Consulting
424 West 49th Street, #3A
New York, NY 10019
Tel: 212-307-1663
Email: kimsabo@aol.com

Dana Fusco, Ph.D.
Department of Teacher Education
York College, CUNY
94-20 Guy R. Brewer Boulevard
Jamaica, NY 11451
Tel: 718-262-2698
Email: fusco@york.cuny.edu

Participatory Evaluation

As an approach to organizational and program assessment, participatory evaluation (PE) involves a collaboration between evaluator and client. Participatory evaluation:

- Increases the likelihood that evaluation results are accurate and relevant
- Ensures that program staff will be motivated and prepared to use evaluation findings
- Builds the program's capacity to continue to design and conduct quality evaluations with diminished reliance on outside assistance
- Stimulates deep thinking about programmatic issues, often leading to refinement in the program itself
- Gives staff new tools for communicating their program to others

ProDES:
A Continuous Improvement System for Juvenile Justice

Philip Harris and Lori Grubstein of the Crime and Justice Research Center describe the "bottom-up" development of ProDES, an outcome-based information system that tracks youth in the juvenile justice system.

ProDES (the Program Development and Evaluation System) is an information system that tracks every court-committed Philadelphia delinquent in the juvenile justice system and measures program outcomes. It is designed to support both the simultaneous continuous improvement of individual programs and the larger system of services to which each program is linked.[1] ProDES provides programs and the system as a whole with a continuous flow of intermediate (changes during the program) and ultimate (recidivism and community adjustment[2]) outcome information. This information is then used for program development, matching youths to programs, and assessment of the program resources available to system-level administrators.

[1] ProDES was developed as a joint venture between the city's Department of Human Services (DHS), Family Court, and the Crime and Justice Research Center.

[2] "Recidivism," which generally means "re-offending," is defined for use in ProDES as a petition to Family Court for a new offense. By "community adjustment" we mean that the youth is attending school regularly or is employed, is getting along well with family members, is not abusing drugs or alcohol, and has not been involved in new delinquent behavior.

The system comprises data collection, analysis, and feedback. ProDES provides four types of information that allow program administrators to make informed changes to program designs, and allow system administrators to assess the value of different programs for different categories of youths and to identify the need for new or expanded program resources. The information consists of:

• Case-specific reports at intake, discharge, and follow-up
• Semi-annual, program-specific reports summarizing aggregated outcome data
• Annual, system-wide reports that describe trends for the entire system
• Topical reports that explore policy issues such as "the female delinquent" and "substance-abusing delinquents"

With this information, programs can study the reasons why some youths leave programs early, explore changes in their incoming client population over time, examine client change on dimensions relevant to future behavior, and develop testable theories on the impact of program improvements.

A Focus on Utilization and Continuous Improvement
We began developing ProDES with a clear bias—that whatever kind of system we created, the information had to be useful to

Goals and Intended Utilization of ProDES

Goal	Information	Intended User Groups	Intended Use
Program Development To facilitate program development through the use of continuous outcome information	• Trends in intake population characteristics • Trends in specific program outcomes • Comparisons to similar programs serving similar clients • Predictive analyses of program outcomes	• Program administrators	• Revision of target population characteristics • Better screening of referrals • Program design modification • Marketing
System Planning To facilitate planning of services through information about program strengths and weaknesses	• Trends in program outcomes • Trends in intake population • Comparisons of similar programs with similar clients • Cost benefit analysis	• Court and Community Services Planning Group – reviews and renews contracts annually • Human Services Monitors – monitor programs regarding compliance with regulations	• Renewal of contracts • Development of requests for proposals • Program accountability for improving poor outcomes • Adjustment of per diem payments
Client Matching To facilitate rational matching of youths to programs	• Program outcomes subdivided by risk of recidivism and case type • Detailed information on program design and structure	• Probation officers • Special Needs Unit Staff • Judges	• Consideration of predictions of outcomes when making program recommendations and decisions

stakeholders for purposes of program and system development. At the same time, we recognized that the information would eventually be used to hold programs accountable for their outcomes. We also recognized the risks inherent in building a data collection system across nearly 100 programs, most of which are operated by private agencies. We could not depend on coercion to build the system; it had to appeal to its user groups.

In planning ProDES, the developmental process had to be "bottom-up" with the private agencies having direct input to both the structure and content of the system. Accordingly, the first phase of the system involved research with 15 agencies that were selected to represent a broad range of programs. Within each program, we conducted evaluability assessments that sought to articulate the goals and objectives of the various programs, their measures of success or failure, and the range of information they routinely collect and use.

The first phase demonstrated that agencies commonly agree on what type of information is important for their programs and the system as a whole. They also agreed about the ideal structure of an information system—one which went beyond simply describing program clients at one point in time and included multiple measures to demonstrate change over time.

Once we identified our key outcome dimensions, we devoted several months to testing standardized scales that would be used to measure change on these key variables. The 15 programs supported the effort by testing the measures on their clients and providing feedback. These same programs served as the core of ProDES when it went into full operation.

A significant part of our job involves facilitated discussions about utilization, helping users develop their own research questions and supporting their use of the results of our analyses. In our utilization design (see box on page 13), each major decision is linked to the appropriate decision maker and to the information that best supports that decision.

Through surveys and an independent evaluation we know that stakeholders' use of ProDES has grown over time. Programs use ProDES to:

- Refine their target population characteristics or screening criteria
- Modify their designs and document their need for additional resources
- Better market their services to the larger system
- Help new managers better understand their strengths and weaknesses

The services system uses ProDES to:

- Review program contracts
- Explore problems with specific programs
- Answer policy-relevant questions (e.g., Why are so many first time offenders being sent to residential facilities? Why are some youths not receiving aftercare services?)
- Plan for future resource demands

The Court and the Probation Department use ProDES to:

- Learn about trends in the delinquent population
- Improve their matching of youths to programs
- Track program use
- Develop plans for future resource needs

Utilization of data proceeded most rapidly among programs, although about three years passed before the use of ProDES data became institutionalized. This adaptation has been facilitated by changes in human services, where there is a growing respect for program evaluation, but working to make evaluation a collaborative venture clearly made a difference.

Developing consistent use in the Court has been more challenging. We developed a working group of Probation Department staff that meets monthly, began conducting training sessions within the department and with the judges, and secured separate funding to work with the Court to develop a tool that will assist judges in matching youths to programs. Support for ProDES is growing steadily as the Court has begun to view its data and staff as supports for developing meaningful goals.

Conclusion

Building an outcome-based information system for an entire juvenile justice system has been, and continues to be, a challenge, even after eight years of operation. We have learned that the strategy of setting program development as our primary goal has worked well. It is a goal that makes sense to all system stakeholders and that enabled us to build working relationships with our primary sources of data and the most critical user group.

We have also learned that face-to-face interactions with users about their outcome data have helped to build the trusting and productive relationships that are necessary to guarantee data utilization. Moreover, these relationships have empowered users to collaborate in our research process, often raising valuable research questions which we can respond to quickly.

Lastly, we have learned the value of both training and effective communication. In order to increase the capacities of users to consume research information we need to make it accessible. This means continually improving our presentations of information as well as training users to become smart consumers of research.

Philip W. Harris
Co-Director
Crime and Justice Research Center
520 North Columbus Boulevard, Suite 600
Philadelphia, PA 19123
Tel: 215-204-5766
Email: pharri01@temple.edu

Lori Grubstein
Former Senior Research Associate
Crime and Justice Research Center
Email: lgrubst@msn.com

Evaluating Citizen Schools

Charlie Schlegel of Citizen Schools explains how their evaluation strategy successfully balances the need to determine program impact with the need for continuous improvement.

Founded in 1995, Citizen Schools operates after school programs in 16 public schools, including 12 in Boston and 4 affiliated sites nationally. Each week, almost 1,000 middle school students—most from low-income households—participate in up to 16 hours of engaging, hands-on learning activities to improve their skills in writing and data analysis as well as their leadership and civic interest. Students also participate in "apprenticeships" taught by adult volunteers, who share their expertise and professionalism through 10-week projects, culminating in a final product or performance.

Citizen Schools Evaluation

Evaluation plays two important roles in the success of our program. First, regular and rigorous assessments of the program quality and impact offer us, as well as policymakers, objective measures of its effect on student achievement and "life trajectory." Second, using the balanced scorecard and other approaches, evaluation generates the performance data and constituent feedback necessary for continuous improvement.

Evaluation at Citizen Schools translates our broadly-stated mission—"to educate kids and strengthen communities"—into four outcome areas: community, leadership, access, and skills (CLAS). Each aspect of the program—the writing and data projects, volunteer-led apprenticeships, homework assistance, and community explorations—are intended to improve students' lives in one or more of the outcome areas. We use four evaluation approaches to assess program impact and analyze data for continuous improvement.

Approach One – Skill Assessments

Citizen Schools assesses students' writing, data, and oral skills at the beginning and end of each school year. Our staff uses this information to identify the specific skill areas in which students need the most help and where they improve over the school year. We use the results and analysis of each skills assessment to make changes in the way our different sites allot time and staff re-

---

Related Resources

The Balanced Scorecard Institute:
www.balancedscorecard.org.

Pane, N., Mulligan, I., Ginsburg, A., & Lauland, A. (1999, February). *A guide to Continuous Improvement Management (CIM) for 21st Century Community Learning Centers.* Washington, DC: U.S. Department of Education. Available at www.ed.gov/offices/OUS/PES/21cent/cim226.pdf (Acrobat file).

---

sources towards each skill area and to adjust the curriculum and structure of our skill-based projects.

Approach Two – Constituent Surveys

Midway through each school year and at the close of the program Citizen Schools asks students, parents, volunteers, school teachers, and staff for feedback through its constituent surveys. The surveys ask about constituents' experience in the program, their assessments of its impact, and suggestions for improvement. We analyze constituent feedback, looking for trends and specific responses that may lead to improvement in the quality of the program and its measures of impact. Over time survey feedback has resulted in programmatic changes, such as additional supports for adult volunteers and different marketing strategies for different constituents.

Approach Three – The Balanced Scorecard

The balanced scorecard is a management system that enables organizations to clarify their vision and strategy and translate them into action. It provides feedback around both internal program processes and external outcomes in order to continuously improve strategic performance and results. (The Balanced Scorecard Institute)

Citizen Schools uses a balanced scorecard to monitor our performance in all areas. Our balanced scorecard includes several program objectives, as well as fundraising, staff management, and national and state policy objectives. We report our progress on each of the balanced scorecard measures to our Board of Directors, staff, and key stakeholders quarterly. Each year the leadership of Citizen Schools establishes the specific objectives and measures of the balanced scorecard for that year based on feedback from staff and other stakeholders within the organization. Staff members collect data for the scorecard and track progress on objectives specific to their departments.

Over time Citizen Schools has found the balanced scorecard to be an effective organizational lever, helping to draw managerial attention and resources to under-resourced initiatives and priorities. For instance, our 2002 balanced scorecard includes "measures" to develop plans for regularly collecting student academic data and to design baseline measures to assess the program's impact on adult volunteers. In short, the balanced scorecard helps keep the scarce resources of the organization focused on the essentials of running high-quality after school programs.

Approach Four – The Impact Study

The Impact Study is a longitudinal, quasi-experimental evaluation conducted by Policy Studies Associates (PSA), an evaluation firm headquartered in Washington D.C. Begun in the fall of 2001, this evaluation uses several methods of data collection and analysis to addresses questions of program implementation, outcomes, and impact on students' school performance and social maturity. This externally-driven evaluation helps us collect the

> beyond basic training

Advice From the Experts on
Nurturing Strong Full Service Schools

The following are excerpts from an evaluation panel at the conference, "Nurturing Strong Full Service Schools: Building Bridges with Communities," that took place on May 20, 2002. It was the fifth in a series of national conferences about full service schools organized by Margot Welch and the Collaborative for Integrated School Services at the Harvard Graduate School of Education. Panelists shared their evaluation findings and lessons learned.

Joy Dryfoos, Moderator
Writer and Researcher, New York, New York
The Collaborative has invited four of the country's leading researchers, each of whom is looking intensively at different components of community schools, particularly after school programs and family resource centers. They have been asked to comment on the state of the art in school-community-partnership programs and address the questions, "What works and what doesn't? How can practitioners approach evaluating their own programs?"

Robert J. Illback, Executive Director
R.E.A.C.H., Louisville, Kentucky
In evaluating the Kentucky Family and Youth Service Centers, we learned the importance of getting people to focus on self-evaluation and program improvement, as opposed to merely gathering data for self-justification or accountability purposes. One of the challenges evaluators face is that in the process of "selling" the initial program concept, large and hard-to-measure goals are often specified to show the importance of the intervention. When program people feel pressure to quickly "prove" that the program works ("because the government is spending all this money"), this can lead to an unrealistic and infeasible evaluation strategy. The net result can be a focus on measures no one thinks are sensitive to what the program is actually doing, with the potential to grossly misrepresent the program's effectiveness. In some ways, you can set yourself up for failure.

Nonetheless, it is important at the outset to consider the information policymakers consider important, because evaluators have a responsibility to funders to report whether the program is moving in the intended direction. But it is also important to stave off the summative questions for as long as possible. Oftentimes we sell programs based on big concepts like, "we want all children to learn at high levels" or "we're going to improve family life or transform communities." Even if we were to accomplish these things, we couldn't prove it had anything to do with the program.

As evaluators, our challenge is to focus program planners on evaluating more proximal outcomes. So, rather than talk about those big issues, let's reframe our focus to consider what we expect will happen in the classroom when a child comes to school, and how we will know whether change has occurred, and how we can measure that reliably.

My tips for designing your evaluation are:

- Invest upfront in understanding the logic of your program. A successful evaluation needs consensus at the beginning about what you're trying to accomplish. Through logic modeling, you can arrive at consensus among various stakeholders.
- Be aware of the political context and what the issues are. Programs are changing and evolving, the climate is always shifting, and legislators and administrators come and go. Recognize that there needs to be a core consensus about what information is being sought from the evaluation—this can change over time.
- Be realistic. Many researchers and psychologists are enamored with the idea that we can have controlled experiments. In community-school partnership evaluations, you need to drop these pretenses. The evaluation is about helping people gather data so that they can achieve understanding.
- Manage information wisely. It's not about the right software program—it's about conceptualizing the information at the outset.
- Have clear questions. We have lots of questions, but we need to hammer down what's doable.
- Look at proximal, not distant, outcomes.
- Keep your eyes on the prize. You have a limited amount of evaluation capital to spend so focus on core questions.

Jean Baldwin Grossman
Private/Public Ventures, Philadelphia, Pennsylvania
Public/Private Ventures and the Manpower Demonstration Research Corporation have just completed an evaluation of the extended service school initiative that began in 1997. Twenty communities that participated in this initiative adopted one of four nationally recognized extended school models: Beacon Schools, Bridge to Success, Community Schools, or West Philadelphia Improvement Corporation.

We found that these after school programs can be put in place quickly and matured over the three years. The programs became much better at figuring out what their core goals were and how to develop activities to meet them. They got better at strategically including children and also became an accepted partner in the schools by the second year.

The after school programs provided developmental supports for children and youth. What mattered for the strength and quality of the program was not the topic or activity or the skill taught; what mattered was the ability of staff to engage and stretch the children.

It takes three or four semesters of participation before you see program effects, especially the academic ones. Staff had to make continuous efforts to attract and hold onto older children, starting with those in the fourth and fifth grades. The choices that programs make affect who shows up. Older children and youth need their own activities that they are interested in.

Children and youth who participated in these programs were likely to show improved school attitudes and behaviors and to stay out of trouble. Children were more likely to handle anger appropriately, to pay attention in class, and to be proud of school. They were also less likely to start skipping school and to start to drink alcohol. However, we cannot know with all certainty whether these effects were due to the program or whether the children who were better behaved and better academically were the ones participating in the first place.

Mark Dynarski, Senior Researcher
Mathematica Policy Research Institute, Princeton, New Jersey
When full-service school programs are funded by short-term grants, an evaluation's results can be fed back to the community for support. The school district is listening to its customers, which are the parents. Community organizations have local boards and are listening to residents. These parents and community groups are the ultimate client who can push for the stability of grant-funded programs. If you don't have a basis for support among the ultimate client, then programs will be short-lived.

The risks in evaluation are that you might not show improvement early on. Then you risk having your client base eroded, or you will be asked to show improvement too early and the same thing will happen. My advice is not to overstate the potential

Related Resources

Kalafat, J., & Illback, R. J. (1999, December). *Evaluation of Kentucky's School-Based Family Resource and Youth Services Center, part 1: Program design, evaluation conceptualization, and implementation evaluation.* Louisville, KY: R.E.A.C.H. of Louisville, Inc. Available at www.reachoflouisville.com/Evaluation%20of%20Family%20Resource%201.htm.

Harvard Family Research Project's Out-of-School Time Evaluation Database. Available at www.gse.harvard.edu/hfrp/projects/afterschool/evaldatabase.html.

Grossman, J. B., et al. (2002, June). *Multiple choices after school: Findings from the Extended-Service Schools Initiative.* Philadelphia: Public/Private Ventures. Available at www.ppv.org/indexfiles/pubsindex.html.

University of Minnesota, College of Education and Human Development, Center for Applied Research and Educational Improvement. *Use of continuous improvement and evaluation in before- and after-school programs: Final report.* (2001). Minneapolis, MN: Author. Available at education.umn.edu/CAREI/Programs/mott/default.html.

gains of the partnership in the work plan. Statements of possible outcomes might sound tremendous.

For example, one health program set a goal of 50% decrease of the infant mortality rate. The only time you see this impact is in developing countries where the water system is so bad that any improved water system reduces diarrhea and other diseases and contributes to dramatic decreases in infant mortality. This is not the case in the U.S. The health program did not meet its goal; it was too large an outcome to go after.

Set realistic and sensible goals, and acknowledge that impacts will be incremental. Don't strive to see the highest thing you've seen in the research. Think about outcomes to be average rather than the extreme.

Heather Weiss, Director
Harvard Family Research Project, Cambridge, Massachusetts
We need to move away from a model where we do an evaluation and if it works and shows success we use it with our funders and if it doesn't we put in a drawer and hope that nobody asks for it. We need to move from evaluation as a one-time effort to evaluation as a continuous learning process. Programs that get ahead of the game will get and use data in an ongoing way to show that they are delivering value to their community over time.

Evaluators and practitioners need to think in terms of continuous improvement and building a *learning system* for the field of full service schools. Our next steps should focus on framing the full service school research and evaluation agenda to be pursued individually and collectively. Evaluators and other stakeholders have to tap and deliver on what people want to know. As we answer questions, new ones open up for inquiry; as we accumulate knowledge, we begin to connect the information to deepen our understanding, and also to identify the gaps that need to be investigated in future research and evaluation. For example, we have to better understand and demonstrate the direct or indirect relationship between after school programs or support services for children and families and educational outcomes.

As part of the learning system, it is important to build a knowledge base that community and school programs can get access to and contribute to. There have been very limited investments in evaluation and knowledge development, and even more limited investments for getting evaluations broadly disseminated so that people can get their hands on them and discuss their implications.

In addition, there's a growing demand for evaluator-practitioner forums that promote cross-sector dialogue and learning. The time has come for funders to invest in virtual and in-person meetings where evaluators talk with practitioners about the evaluation findings and their implications for the continuous improvement of programs and for learning and field building.

This is what it means to build a learning system for a field. I encourage the dialogue to continue about what this system should look like for the field of full service schools.

Past articles on school-linked services can be found in The Evaluation Exchange *Vol. III, No. 2, 1997. Archived issues are available at www.gse.harvard.edu/hfrp/eval/archives.html.*

Risk
continued from page 9

says, pointing to the example of Thomas Edison discovering thousands of ways *not* to invent the incandescent electric lightbulb.

In explaining how we assess risk, I start with a simple example. Let's say the desired outcome is for me to lose 20 pounds. Three types of risk come into play.

There's *idea risk*. This represents the risk associated with the idea's track record and the logic connecting the activities to the desired result. I found my favorite example of idea risk in *The New York Times* in summer 2001. Thieves stole art from a Jewish museum in New York and pledged to hold the objects until there was peace in the Middle East. Talk about your poorly formed theory of change. It's just the opposite with my weight-loss example. For me, a low-fat diet with regular aerobic exercise carries almost no idea risk.

That brings us to *implementation risk*. Bad ideas can be well-implemented and vice versa. My diet-and-exercise solution fails because of botched implementation—I haven't been able to stick with the regimen. In philanthropic terms, I wouldn't give me a grant.

Untangling idea risk and implementation risk helps our staff integrate evaluation into grant development. When do we need more conceptual work? When do we need technical assistance to ensure effective execution?

Then there's *evidence risk*. In my weight-loss example, there's almost no evidence risk. Not only will the evidence be unambiguous evidence, it's quite depressing.

But with most grants, evidence is dicey. It's hard to agree on evidence standards. Then detecting effects is inexact. We risk not knowing one way or the other. Deciding when it's important to diminish evidence risk helps with front-end evaluation planning. These discussions also help us reach consensus on investments that merit scarce evaluation dollars.

We've got a growing list of factors that can affect all three types of risk, and we have added a section on risk in the grant summaries we write for our board (see box).

With this prospective analysis, we don't pretend to predict the future. We need a clear account of our original expectations to honestly compare purported risks against hazards actually encountered.

We are currently working with trustees to set tolerances for idea risk. We want agreement on how much we're willing to invest in approaches that that tilt toward the tried-and-true versus experimental approaches that ensure higher rates of failure, but promise greater learning opportunities.

John Bare, Ph.D.
Director of Program Development and Evaluation
John S. and James L. Knight Foundation
One Biscayne Tower, Suite 3800
2 South Biscayne Boulevard
Miami, FL 33131-1803
Tel: 305-908-2600
Email: bare@knightfdn.org

Risk Factors

Below are factors that can affect how risky an investment an initiative is for a funder. In grant development, it's best to focus on the factors that need to be addressed to increase the likelihood of success for the investment.

- Environmental turbulence
- Needs of target population
- Time horizon
- Implementer's history
- Size of the investment
- Clarity of logic model
- Sustainability
- Degree of controversy
- Complexity
- Leadership
- Novelty of the idea
- Nature of capacity needed
- Visibility of the effort
- Partners' relationships
- Number/nature of foundation investors
- Degree of consensus on the local advisory committee
- Degree of politicization
- Potential for negative side effects
- Rifle-shot approach vs. natural variation (i.e., are alternatives funded?)
- Depth of knowledge base
- Hidden assumptions

Evaluating Citizen Schools
continued from page 15

necessary outcomes data so our staff can focus on collecting data for continuous improvement.

The relationship between internal efforts to use data for continuous improvement and externally-generated data for proving impact is complementary. Citizen Schools and PSA both collect data about students' school performance and social maturity—data that we use to measure the impact of our program as well as inform our decisions about quality improvements in our services.

Charlie Schlegel
Director of Research and Evaluation
Citizen Schools
Museum Wharf
308 Congress Street
Boston, MA 02210
Tel: 617-695-2300
Email: charlieschlegel@citizenschools.org

Increase Equity

feedback on performance, public recognition, meetings with policymakers and senior managers, and increases in flexibility. Financial incentives may include bonuses, increases in budget levels, and award or renewal of discretionary grants and contracts.

Efforts to enhance equity may draw inspiration from the success of others. For example, the Veteran's Health Administration and the U.S. Department of Transportation have used performance agreements to focus executives' attention on agency performance goals.[5] The County Board of Arlington, Virginia, used frequent public reporting and meetings with the County Manager to stimulate the hiring of minorities and women. At national, state, and local levels, enforcement of the Voting Rights Act of 1965 has changed the incentive systems of elected officials and helped create greater equity in public services.

Joseph S. Wholey
Professor of Public Administration
University of Southern California
2103 21st Road North
Arlington, VA 22201
Tel: 703-524-0022
Email: wholey@usc.edu

[5] U. S. General Accounting Office. (2000). *Managing for results: Emerging benefits of selected agencies' use of performance agreements.* Washington, DC: Author.

Reflective Assessments

such as an extended meeting or retreat, in which participants can reflect and react to the report. This meeting, and subsequent discussions, allow participants to focus on their core concerns: Are we on the right track? How do others view our effort? Are we encountering barriers that we didn't anticipate? Should we make any changes in our approach? The assessment report does not sit on a shelf; rather it becomes a critical tool for charting the next steps.

While not the right tool for all efforts, at all times, reflective assessments fill an important niche in the evaluation landscape. A particularly important attribute has been the *responsiveness* of the process; a reflective assessment takes weeks, not years, to produce usable information.

Sharon Edwards
President
Cornerstone Consulting Group
One Greenway Plaza, Suite 550
Houston, TX 77046
Email: sedwards@cornerstone.to

Ira Cutler
Vice President
Cornerstone Consulting Group
181 Westchester Avenue
P.O. Box 791
Port Chester, NY 10573
Email: icutler@cornerstone.to

Results Accountability

sign. It is that kind of "lesson learned" that may be helpful to other states and that needs to be explored and documented.

Assuming Accountability and Supporting Improvement

One of the biggest problems that a formal accountability system can have is a process that moves directly from collecting the numbers to follow-up actions, without sufficient intervening analysis about why a certain level of performance happened and what the next steps should be. Poor performance might be the result of:

- The wrong strategy
- The right strategy, poorly implemented or inadequately funded
- Another partner failing to fulfill its responsibility
- Some outside factor (e.g., the economy, or a population with more challenges than anticipated)

Knowing which reason is the correct one is essential to deciding on next steps. The box on page 6 lists possible follow-up actions—both "carrots" and "sticks"—for different levels of performance. They are based on choices states and localities are currently using, and can be applied at the individual and organizational levels.

It is immensely challenging to decide on and implement follow-up actions that address the correct problem, hold people responsible for the appropriate level of performance (e.g., don't penalize individuals for performance on results that are much broader than they can hope to affect and vice-versa), increase trust, improve performance, and satisfy the public and the media that progress is being made. Making the right decisions about who should be accountable to whom for what is crucial. Mistakes here can doom accountability systems by raising fears and resistance among those who are responsible for improvements.

Most accountability systems are just beginning to complete the cycle of collecting data, analyzing the reasons behind performance, and taking action to improve. Just as experts and practitioners have produced hundreds of reports and guidebooks analyzing different definitions of data variables, we need commensurate attention to who should do what based on the data. Only then will results accountability systems fulfill their potential of improving outcomes for children and families.

Sara Watson
Program Officer, Education
The Pew Charitable Trusts
2005 Market Street, Suite 1700
Philadelphia, PA 19103
Tel: 215-575-9050
Email: swatson@pewtrusts.com

The opinions expressed in this article are those of the author and do not necessarily reflect the views of the Pew Charitable Trusts. Past articles on results accountability can be found in The Evaluation Exchange *Vol. IV, No. 1, 1998, Vol. II, No. 3, 1996, and Vol. II, No. 1, 1996. Archived issues are available at www.gse. harvard.edu/hfrp/eval/archives.html.*

Bare, J. (2002). **Try my elixir**. *Foundation News & Commentary. 43*(1), 22-24. Acknowledging that non-evaluators tend not to care much about how evaluators do their jobs, Bare presents 10 ways to communicate the value of evaluation to grantmaking colleagues.

The Annie E. Casey Foundation. ***Children at risk: State trends 1990-2000.*** (2002, March). Baltimore, MD: Author. A Population Reference Bureau/Kids Count report based on the Census 2000 Supplementary Survey, which finds that, while conditions for children improved overall in the last decade, the child poverty rate remained practically unchanged. **www.aecf.org/kidscount/c2ss**

The Robert Wood Johnson Foundation **Pursuing Perfection initiative** is "intended to help physician organizations and hospitals dramatically improve patient outcomes by pursuing perfection in all of their major care processes." **www.ihi.org/pursuingperfection**

Leichter, H. M., & Tryens, J. (2002, September). ***Achieving better health outcomes: The Oregon benchmark experience.*** New York: Milbank Memorial Fund. This report examines and assesses the history of benchmarking for health by the Oregon Progress Board. **www.milbank.org**

Mosteller, F., & Boruch, R. (Eds.). (2002). ***Evidence matters: Randomized trials in education research***. Washington, DC: The Brookings Institution Press. This book explores the history and current status of research on practices in education, particularly the need for credible studies of the effects of education programs. The contributors focus on randomized field trials, but also consider other approaches to generating empirical evidence, such as sample surveys, narrative studies, and exploratory research. **www.brook.edu/dybdocroot/savingsforthepoor/evidence_matters.htm**

The Pew Partnership for Civic Change's **Pew Civic Entrepreneur Initiative** is a nationwide project to equip a diverse group of citizens, or "civic entrepreneurs," to play a greater role in decision making in their communities. It includes a LeadershipPlenty training curriculum, which has an evaluation component. **www.pew-partnership.org/pcei/pcei.html**

The Results and Performance Accountability Implementation Guide, developed by Mark Friedman, is an Internet-based guide for individuals and groups looking to implement some form of results or performance accountability in their organization or community. **www.raguide.org**

New Resources From HFRP

A **new teaching case** on the standards movement in education is available on our FINE (Family Involvement Network of Educators) website—***Defining "Fine": Communicating Academic Progress to Parents***. Read it at **www.gse.harvard.edu/hfrp/projects/fine/resources/teaching-case/progress.html**.

We have just published **two new research briefs** in our Issues and Opportunities in Out-of-School Time Evaluation series.

• Brief number 3, published in July 2002, entitled ***Documenting Progress and Demonstrating Results: Evaluating Small-Scale Out-of-School Time Programs***, provides small out-of-school time programs with the techniques, tools, and strategies they need to improve their programs and track their effectiveness over time.

• Brief number 4, published in August 2002, entitled ***Beyond the Head Count: Evaluating Family Involvement in Out-of-School Time***, offers an overview of how out-of-school time programs involve families and how they can evaluate family involvement.

The entire series of out-of-school time research briefs is available on our website at **www.gse.harvard.edu/hfrp/projects/afterschool/resources.html#issues** or email us at **hfrp_pubs@gse.harvard.edu** to order hard copies.

contact us @