



# the evaluation exchange

A PERIODICAL ON EMERGING STRATEGIES IN EVALUATING CHILD AND FAMILY SERVICES

HARVARD FAMILY RESEARCH PROJECT ■ HARVARD GRADUATE SCHOOL OF EDUCATION ■ VOLUME XI NUMBER 2 ■ SUMMER 2005

## from the director's desk



Heather Weiss

A discipline's methodology, broadly defined, is its body of methods, approaches, rules, and techniques. Because evaluators come from diverse backgrounds, and because evaluation has become increasingly interdisciplinary, bringing points of view from communications, psychology, public health, and other arenas to bear, we enjoy a richness and variety in our methodology that is unmatched in many other disciplines.

While evaluation methodology is grounded firmly in traditional social science approaches, evaluation is more than the application of social science methods to study social problems. We make judgments and facilitate decisions that require us to regularly engage with programs and stakeholders and require us to integrate empirical evidence with standards and values to reach evaluative conclusions. As such, our methodology has evolved to embrace novel ideas and tactics that help us perform this complex work.

A fundamental benefit of having such a diverse methodology is that with every evaluation we have *choices*. The methods used for one evaluation can differ greatly from the next, depending on what we are trying to accomplish or answer and the theories and preferences we bring to the table. For example, if our priority is making sure our audiences use our work, we might choose a utilization-focused approach. If our priority is answering as unequivocally as possible "what works," we may choose a randomized trial. If our priority is engaging stakeholders and building evaluation capacity, we may choose an empowerment or participatory approach.

Better yet, our choices are even broader because we can "mix and match" methodological designs (i.e., experimental, quasi-experimental, non-experimental) with different evaluation approaches (e.g., theory-based, goal free, constructivist), qualitative and quan-

titative methods, and useful tools and techniques (e.g., logic models, strategic planning, etc.). The possibilities are exciting, always growing, and help to keep our work fresh and engaging.

This is our third issue of *The Evaluation Exchange* devoted entirely to the theme of methodology, though with every issue we try to identify new methodological choices, the instructive ways in which people have applied or combined different methods, and emerging methodological trends. For example, lately we have seen "theories of change" gain almost buzzword status in our field, and have featured this concept in several recent issues, including this one.

Other topics explored here include the role that evaluation theory plays in our methodological choices, the proliferation of outcome models during the recent movement toward more accountability for social programs, how to give the connection between evaluation and learning more than just lip service, and developments in the never ending search to ensure our work avoids being forgotten on dusty bookshelves (or hard drives) and instead gets the attention we are so certain it deserves. We also highlight cluster evaluation and retrospective pretests as part of our methodological repertoire, and have several articles on the relationship between evaluation methodology, training, and minority populations.

Because evaluation methodology is constantly in flux, we will revisit this theme frequently in future issues. We will continue to keep our finger on the pulse of emerging trends and innovations, and as always, invite you to share your own observations about new developments or new methodological waters you've charted.

## in this issue

### *Evaluation Methodology*

#### Theory & Practice

Evaluation theory 2

#### Promising Practices

Eight outcome models 4

Choosing measures 6

Capacity assessment 7

#### Ask the Expert

Strategic learning 8

Randomized trials 9

Theory of change 12

#### Questions & Answers

A conversation with Gary Henry 10

#### Evaluations to Watch

Complex initiatives 13

Cluster evaluation 14

#### Beyond Basic Training

Multicultural competency 15

Minority undergraduates 16

Evaluators of color 17

#### Spotlight

Retrospective pretest 18

#### New & Noteworthy

20

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

the

# evaluation exchange



**Founder & Director**

Heather B. Weiss, Ed.D.

**Managing Editor**

Julia Coffman

**Contributing Editor**

Erin Harris

**HFRP Contributors**

Heather B. Weiss, Ed.D.  
Julia Coffman, Erin Harris  
Abby Weiss

**Publications/Communications  
Manager**

Stacey Miller

**Publications Editor**

Tezeta Tulloch

©2005 President &  
Fellows of Harvard College

Published by  
Harvard Family Research Project  
Harvard Graduate School of Education

All rights reserved. This periodical  
may not be reproduced in whole or  
in part without written permission  
from the publisher.

Harvard Family Research Project  
gratefully acknowledges the support  
of the Annie E. Casey Foundation,  
the Marguerite Casey Foundation,  
the Ewing Marion Kauffman Founda-  
tion, the W. K. Kellogg Foundation, the  
John S. and James L. Knight Foundation,  
and the C. S. Mott Foundation.

The contents of this publication  
are solely the responsibility of  
Harvard Family Research Project  
and do not necessarily reflect  
the view of our funders.

*The Evaluation Exchange* accepts  
query letters for proposed articles.  
See our query guidelines on our website  
at [www.gse.harvard.edu/hfrp/eval/submit.html](http://www.gse.harvard.edu/hfrp/eval/submit.html)

To request a free subscription to  
*The Evaluation Exchange*, email us at  
[hfrp\\_pubs@gse.harvard.edu](mailto:hfrp_pubs@gse.harvard.edu) or  
call 617-496-4304.

## > theory & practice

### Evaluation Theory or What Are Evaluation Methods for?

*Mel Mark, professor of psychology at the Pennsylvania State University and president-elect of the American Evaluation Association, discusses why theory is important to evaluation practice.*

**H**ey, this issue of *The Evaluation Exchange* focuses on methods, including recent methodological developments. What's this piece on evaluation theory doing here? Was there some kind of mix-up?

No, it's not a mistake. Although evaluation theory<sup>1</sup> serves several purposes, perhaps it functions most importantly as a guide to practice. Learning the latest methodological advance—whether it's some new statistical adjustment for selection bias or the most recent technique to facilitate stakeholder dialogue—without knowing the relevant theory is a bit like learning what to do without knowing why or when.

What you risk is the equivalent of becoming really skilled at tuning your car's engine without thinking about whether your transportation needs involve going across town, overseas, or to the top of a skyscraper. Will Shadish, Tom Cook, and Laura Leviton make the same point using a military metaphor: "Evaluation theories are like military strategies and tactics; methods are like military weapons and logistics," they say. "The good commander needs to know strategy and tactics to deploy weapons properly or to organize logistics in different situations. The good evaluator needs theories for the same reasons in choosing and employing methods."<sup>2</sup>

The reasons to learn about evaluation theory go beyond the strategy/tactic or why/how distinction, however. Evaluation theory does more than help us make good judgments about what kind of methods to use, under what circumstances, and toward what forms of evaluation influence.

First, evaluation theories are a way of consolidating lessons learned, that is, of synthesizing prior experience. Carol Weiss' work can help evaluators develop a more sophisticated and nuanced understanding of the way organizations make decisions and may be influenced by evaluation findings.<sup>3</sup> Theories enable us to learn from the experience of others (as the saying goes, we don't live long enough to learn everything from our own mistakes). George Madaus, Michael Scriven, and Daniel Stufflebeam had this function of evaluation theory in mind when they said that evaluators who are unknowledgeable about theory are "doomed to repeat past mistakes and, equally debilitating, will fail to sustain and build on past successes."<sup>4</sup>

Second, comparing evaluation theories is a useful way of identifying and better understanding the key areas of debate within the field. Comparative study of evaluation theory likewise helps crystallize what the unsettled issues are for practice. When we read the classic exchange between Michael Patton and Carol Weiss,<sup>5</sup> for example, we learn about very different perspectives on what evaluation use can or should look like.

A third reason for studying evaluation theory is that theory should be an important part of our identities as evaluators, both individually and collectively. If we think of ourselves in terms of our methodological skills, what is it that differentiates us from many other people with equal (or even superior) methodological expertise? Evaluation theory. Evaluation theory, as Will Shadish said in his presidential address to the American Evaluation Association, is "who we are."<sup>6</sup> But people come to evaluation through quite

1. Although meaningful distinctions could perhaps be made, here I am treating *evaluation theory* as equivalent to *evaluation model* and to the way the term *evaluation approach* is sometimes used.
2. Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
3. For a recent overview, see Weiss, C. H. (2004). Rooting for evaluation: A Cliff Notes version of my work. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 12–65). Thousand Oaks, CA: Sage.
4. Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (1983). *Evaluation models*. Boston: Kluwer-Nijhoff.
5. See the papers by Weiss and Patton in Vol. 9, No. 1 (1988) of *Evaluation Practice*, reprinted in M. Alkin (Ed.). (1990). *Debates on evaluation*. Newbury Park, CA: Sage.
6. Shadish, W. (1998). Presidential address: Evaluation theory is who we are. *American Journal of Evaluation*,

varied pathways, many of which don't involve explicit training in evaluation. That there are myriad pathways into evaluation is, of course, a source of great strength for the field, bringing diversity of skills, opinions, knowledge sets, and so on.

Despite the positive consequences of the various ways that people enter the field, this diversity also reinforces the importance of studying evaluation theories. Methods are important, but, again, they need to be chosen in the service of some larger end. Theory helps us figure out where an evaluation should be going and why—and, not trivially, what it is to be an evaluator.

Of course, knowing something about evaluation theory doesn't mean that choices about methods can be made automatically. Indeed, lauded evaluation theorist Ernie House notes that while theorists typically have a high profile, “practitioners lament that [theorists'] ideas are far too theoretical, too impractical. Practitioners have to do the project work tomorrow, not jawbone fruitlessly forever.”<sup>7</sup> Especially for newer theoretical work, the translation into practice may not be clear—and sometimes not even feasible. Even evaluation theory that has withstood the test of time doesn't automatically translate into some cookbook or paint-by-number approach to evaluation practice.

More knowledge about evaluation theory can, especially at first, actually make methods choices harder. Why? Because many evaluation theories take quite different stances about what kind of uses evaluation should focus on, and about how evaluation should be done to achieve those uses. For example, to think about Donald Campbell<sup>8</sup> as an evaluation theorist is to highlight (a) the possibility of major choice points in the road, such as decisions about whether or not to implement some new program; (b) the way decisions about such things often depend largely on the program's potential effects (e.g., does universal pre-K lead to better school readiness and other desirable outcomes?); and (c) the benefits of either randomized experiments or the best-available quasi-experimental data for assessing program effects.

In contrast, when we think about Joseph Wholey<sup>9</sup> as an evaluation theorist, we focus on a very different way that evaluation can contribute: through developing performance-measurement systems that program administrators can use to improve their ongoing decision making. These performance measurement systems can help managers identify problem areas and also provide them with good-enough feedback about the apparent consequences of decisions.

Choosing among these and the many other perspectives available in evaluation theories may seem daunting, especially at first. But it's better to learn to face the choices than to have them made implicitly by some accident of one's methodological training. In addition, theories themselves can help in the choosing. Some evaluation theories have huge “default options.” These theories may not exactly say “one size fits all,” but they certainly suggest

that one size fits darn near all. Indeed, one of the dangers for those starting to learn about evaluation theory is becoming a true believer in one of the first theories they encounter. When this happens, the new disciple may act like his or her preferred theory fits all circumstances. Perhaps the most effective antidote to this problem is to be sure to learn about several evaluation theories that take fairly different stances. Metaphorically, we probably need to be multilingual: No single evaluation theory should be “spoken” in all the varying contexts we will encounter.

However, most, if not all, evaluation theories are contingent; that is, they prescribe (or at least are open to) quite different approaches under different circumstances. As it turns out, there even exist theories that suggest very different bases for contingent decision making. Put differently, there are theories that differ significantly on reasons for deciding to use one evaluation design and not another.

These lead us to think about different “drivers” of contingent decision making. For example, Michael Patton's well-known *Utilization-Focused Evaluation* tells us to be contingent based on *intended use by intended users*. Almost any method may be appropriate, if it is likely to help intended users make the intended use.<sup>10</sup> Alternatively, in a recent book, Huey-Tsyh Chen joins others who suggest that the choices made in evaluation should be driven by *program stage*.<sup>11</sup> Evaluation purposes and methods for a new program, according to Chen, would typically be different from those for a mature program. Gary Henry, George Julnes, and I<sup>12</sup> have suggested that choices among alternative evaluation purposes and methods should be driven by a kind of analytic assessment of each one's *likely contribution to social betterment*.<sup>13</sup>

It can help to be familiar with any one of these fundamentally contingent evaluation theories. And, as is true of evaluation theories in general, one or another may fit better, depending on the specific context. Nevertheless, the ideal would probably be to be multilingual even in terms of these contingent evaluation theories. For instance, sometimes intended use may be the perfect driver of contingent decision making, but in other cases decision making may be so distributed across multiple parties that it isn't feasible to identify specific intended users: Even the contingencies are contingent. Evaluation theories are an aid to thoughtful judgment—not a dispensation from it. But as an aid to thoughtful choices about methods, evaluation theories are indispensable.

#### Mel Mark, Ph.D.

*Professor of Psychology, The Pennsylvania State University, Department of Psychology, 407 Moore Hall, University Park, PA 16802. Tel: 814-863-1755. Email: m5m@psu.edu*

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

11. Chen, H.-T. (2005). *Practical program evaluation: Assessing and improve planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.

12. Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.

13. Each of these three theories also addresses other factors that help drive contingent thinking about evaluation. In addition, at least in certain places, there is more overlap among these models than this brief summary suggests. Nevertheless, they do in general focus on different drivers of contingent decision making, as noted.

19(1), 1–19.

7. House, E. R. (2003). Evaluation theory. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 9–14). Boston: Kluwer Academic.

8. For an overview, see the chapter on Campbell in the book cited in footnote 2.

9. See, e.g., Wholey, J. S. (2003). Improving performance and accountability: Responding to emerging management challenges. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium*. Mahwah, NJ: Lawrence Erlbaum.

## Eight Outcome Models

Robert Penna and William Phillips from the Rensselaerville Institute's Center for Outcomes describe eight models for applying outcome-based thinking.<sup>1</sup>

Over the past decade, *outcomes* has moved from being just a buzzword in the nonprofit, government, and foundation worlds to becoming a full-fledged movement. As the outcomes movement and outcome-based decision making have grown, many models or frameworks for applying this thinking have emerged. While evaluators and practitioners have benefited greatly from the development of various tools to guide outcomes thinking, understanding the unique advantages of each model and how to select the right one is challenging for many.

Responding to this challenge, the Rensselaerville Institute's Center for Outcomes published in 2004, *Outcome Frameworks: An Overview for Practitioners*, a book that offers insights into which model might be appropriate to the particular needs of a program at a given point in time. It captures what the outcomes movement means, where it came from, the major models now in use, and the movement's probable future. The models described in *Outcome Frameworks* fall into three main categories: program planning and management, program and resource alignment, and program reporting. In addition, most models can be used as an evaluation tool.

### Program Planning/Management Tools

Program planning or management tools are outcome models that assist in an effort's proposal, funding, and implementation phases. They illustrate the logic, theory of change, and anticipated flow of an intervention, providing markers against which both incremental and ultimate progress may be measured. Examples include the following models:

**Model 1: The Logic Model.** Logic models, the most widely used of these models, provide a graphic overview of a program, outlining the outcomes to be accomplished along with how they are to be achieved and for what groups.<sup>2</sup> A logic model generally includes the target group, the resources to be used, activities, and objectives. Best used for describing a program in the broadest strokes, it can be an extremely useful tool, particularly at the earliest stages of a project.<sup>3</sup>

**Model 2: Outcome Funding Framework.** This model stresses key shifts in the thinking that traditionally has influenced human service programs. It encourages funders to think like investors<sup>4</sup> and

encourages programs to shift from emphasizing *service activities*, to focusing on *performance targets*, defined in terms of client changes gained. The model also uses milestones, or sequential steps toward achieving ultimate targets, to allow for ongoing assessment and mid-course program corrections.

**Model 3: Results-Based Accountability (RBA).** This model starts with the desired ends and works backward toward the means to achieve them. RBA first describes what a desired result would look like, then defines that result in measurable terms, and, finally, uses those measures to gauge success or failure. RBA asks and answers three basic questions: What do we want? How will we recognize it? What will it take to get there? This model distinguishes between population accountability and program accountability. Its inclusion of the *crosswalk*, a tool for matching RBA with other outcome models, is a unique and useful aspect of the framework.<sup>5</sup>

**Model 4: Targeting Outcomes of Programs (TOP).** This model is based on a hierarchy of sequential steps in planning, implementing, and evaluating programs. It helps answer four basic questions: Why have a program? How should it be conducted? Has the program design been implemented? What are the benefits delivered?<sup>6</sup>

### Program and Resource Alignment Tools

Program and resource alignment tools ensure that resources and effort are expended in support of organizational goals; one such is the Balanced Scorecard, explained below:

**Model 5: Balanced Scorecard.** Initially designed as a corporate management framework, the balanced scorecard synthesizes multiple measures, reflecting a range of processes, and links them to a consistent and mutually reinforcing whole.<sup>7</sup> The model's use of a resource or target matrix makes it particularly well suited to organizational alignment.

### Program Reporting Tools

Program reporting tools allow organizations to capture and communicate the fullness of the results they have achieved. They include the following:

**Model 6: Scales and Ladders.** This model offers a matrix-based system popularly associated with the implementation of the Results Oriented Management and Accountability system within community-service-block grant-funded programs.<sup>8</sup> The model's essential concept centers on a series of anchored scales and their placement within a matrix that describes different states or conditions of client status along a continuum.

1. The Rensselaerville Institute is a national nonprofit specializing in outcome-tool creation and assisting government, foundations, and nonprofits to put them into practice. Known as "the think tank with muddy boots," the Institute develops new outcomes approaches and tools through its Center for Outcomes, and then applies them through partnerships and independent initiatives. [www.rinstitute.org](http://www.rinstitute.org)

2. Kirkpatrick, S. (2001). *The program logic model: What, why and how?* Retrieved June 8, 2005, from <http://www.charityvillage.com/cv/research/rstrat3.html>

3. MacNamara, C. (2000). *Guidelines and framework for designing basic logic model.* Retrieved June 8, 2005, from [http://www.managementhelp.org/np\\_progs/np\\_mod/org\\_frm.htm](http://www.managementhelp.org/np_progs/np_mod/org_frm.htm)

4. Williams, H. S., Webb, A. Y., & Phillips, W. J. (1991). *Outcome funding: A new approach to targeted grantmaking.* Rensselaerville, NY: The Rensselaerville Institute.

5. Friedman, M. (2001). *The results accountability implementation guide.* Retrieved June 8, 2005, from <http://www.raguide.org>

6. Bennett, C., & Rockwell, K. (1995). *Targeting outcomes of programs (TOP): An integrated approach to planning and evaluation.* Washington, DC: Cooperative State Research, Education, and Extension Service.

7. Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action.* Boston: Harvard Business School Press.

8. Learn more about it at [www.roma1.org](http://www.roma1.org).

*Model 7: Results Mapping.* Results mapping is an outcome-based evaluation tool that systematically captures anecdotal evidence and uses the information to present a results-based conclusion.<sup>9</sup>

*Model 8: Program Results Story.* Currently under development, this approach applies the power of the story format to capture organizations' achievements and present them in a results-based format.<sup>10</sup>

9. Pacific Institute for Research and Evaluation. [www.pire.org](http://www.pire.org)

10. The program results story is not included in *Outcome Frameworks*. Contact the authors for more information on this approach.

**Robert M. Penna, Ph.D.**

Senior Consultant. Tel: 518-438-7502. Email: [rmpc52@aol.com](mailto:rmpc52@aol.com)

**William J. Phillips**

Director. Tel: 518-797-3783. Email: [bphillips@rinstitute.org](mailto:bphillips@rinstitute.org)

Center for Outcomes, The Rensselaerville Institute, 63 Huyck Road, Rensselaerville, NY 12147

## OVERVIEW OF EIGHT OUTCOME MODELS

	Model	Description	Key Concepts	Strong Points	Well Suited For
Program Planning/Management	<b>1. Logic Model</b>	Diagrammatic representation of a program, showing what it is supposed to do, with whom, and why	Inputs, outputs, outcomes; arrows show relationships between elements in the model	Easy to use; provides easily understood representation of program's theory of change	Program overview; presentations; program and evaluation planning
	<b>2. Outcome Funding Framework</b>	Key management focus on the achievement of specific, sequential results for customers of services; emphasis on results, not activity	Investor return, results, customers, milestones, performance targets, outcome statement	Highly disciplined approach that serves both program investors and implementers; Web-based software has strengthened usability	Government and philanthropic grantmaking; program and organization management
	<b>3. Results-Based Accountability</b>	Real-time approach that describes what desired results look like, defines results in measurable terms, and uses measures to drive action plans for improvement	Results, experience, indicators, baselines, strategy, action plan and budget, accountability	Thorough system for planning community- change efforts and improvements in program, agency, or system performance; uses lay language and provides direct link to budgeting; useful for integrating different outcome systems	Project planning and start-up; development of community report cards; program/agency improvement plans and budgets; grantmaking and evaluation design
	<b>4. Targeting Outcomes of Programs</b>	Tracking progress toward achievement targets; evaluating degree to which programs impact targeted conditions	Knowledge, attitude, skills, and aspiration; process, outcome, and impact evaluation	Fairly easy to use; helps integrate program development and evaluation; implementers and managers can use same concepts	Program design and evaluation
Program and Resource Alignment	<b>5. Balanced Scorecard</b>	Business-based model designed to provide integrated management and accounting for multiple variables impacting organization performance by connecting them to a set of performance indicators	Strategy, alignment, short- and long-term objectives; financial and nonfinancial measures; lagging and leading indicators; performance measures and drivers; internal and external indices of success	Allows for a graphic assessment of the degree to which an organization's resources and efforts support its goals	Monitoring either a single program with several associated initiatives or multiple programs within an organization; analyzing alignment of resources and initiatives to strategic targets
Program Reporting	<b>6. Scales and Ladders</b>	Graphic tool that centers around a series of scales and their placement within a matrix designed to illustrate progress along a continuum of stages	Scales; mutually exclusive, multiple, and floating indicators	Places a client, community, or program on a continuum; shows incremental and relative progress, stabilization, or decline; individual data together tell a complete story; behaviorally anchored description of levels of change	Demonstration of aggregate progress; measuring concepts that are not easily quantified
	<b>7. Results Mapping</b>	Outcome-based evaluation tool designed to systematically capture otherwise nonquantifiable anecdotal evidence	Causal and synchronistic attribution; levels and milestones	Way to systemize, standardize, gather, and utilize lessons embedded in anecdotal information	Turning anecdotal information into a useful tool for program presentation, evaluation, and assessment
	<b>8. Program Results Story</b>	Uses stories to capture organizations' achievements and present them in a results-based format	Results, stories, anecdotal evidence	Easily understood approach for presenting results; brings outcomes to human interest level; captures and conveys richness of information	Presenting program and results to multiple audiences

## Evaluation and the Sacred Bundle

*John Bare of the Arthur M. Blank Family Foundation explains how nonprofits can learn about setting evaluation priorities based on storytelling and “sacred bundles.”*

One reason nonprofits struggle with evaluation is that they are usually trying to figure out what someone else—a funder, a government agency, or some expert—wants them to measure. When I talk to nonprofit staff, I ask them to tell me what *they*—not anyone else—need to track while a program is unfolding in order to observe what’s working well or to know when adjustments are needed. This utilization-focused approach creates a bridge from evaluation to storytelling. In the process of figuring out what information they need to track to manage their effectiveness, nonprofit staff realizes that the same information will fuel their own stories.

### New Resources From HFRP

The Family Involvement Network of Educators (FINE) has several new resources to the FINE website ([www.finenetwork.org](http://www.finenetwork.org)):

- The research digest, *Learning by Listening: A Longitudinal Study of Family Literacy*, examines the reading concepts held by urban families and how home reading practices intersect with school literacy practices.
- The University of Utah course **Family Strengthening Interventions: Evidence-Based Practices** looks at evidence-based family strengthening interventions.
- The handouts and slides from the **Research and Evaluation of Family Involvement in Education: What Lies Ahead?** panel session at the 2005 annual conference of the American Educational Research Association are available.
- Also available are the workshop slides from **Making the Case for Parental Involvement and Engagement: Part I: Parental, Family, School, and Community Partnerships Make a Difference**.

[www.gse.harvard.edu/hfrp/projects/fine/announcements/05jun.html](http://www.gse.harvard.edu/hfrp/projects/fine/announcements/05jun.html)

HFRP recently contributed to a publication from Collaborative Communications Group entitled *Moving Towards Success: Framework for After-School Programs*. It is a working document designed to help improve after school programs, developed by a committee of preeminent researchers, evaluators, and program experts. It describes a theory of change approach to help guide the thinking and implementation of program goals and elements, outlines a broad range of desired participant outcomes to consider, and discusses conditions most suitable for achieving positive results that meet the needs of the after school participants. [www.publicengagement.com/Framework](http://www.publicengagement.com/Framework)

### The Sacred Bundle<sup>1</sup>

Some indigenous tribes in North America, especially nomadic tribes, preserved their culture by literally holding on to parts of their history. An elder would be trusted to keep a rock, feather, or other small totem representing an important moment from a tribe’s history. The elder would keep these items in a pouch—a sacred bundle. When the tribe gathered around the evening campfire, the trusted member would retrieve items from the sacred bundle, telling the story associated with each. Over time, everyone in the tribe came to know the same stories.

Dr. Jane Goodall, who has committed her life to improving the situation of endangered species such as chimpanzees and great apes, carries her own sacred bundle. In it she keeps a feather from a peregrine falcon, retrieved from a former mine in Ontario. The mining had left the land as barren as a moonscape. But, eventually, environmentalists reclaimed the land and habitats returned. Falcons are there again, and Goodall carries with her a feather from one of those falcons.

Another item in Goodall’s pouch is a child’s change purse. Once, on a TV show, Goodall featured a chimpanzee raised in an isolation room made of clear plastic. The animal was being used as part of a study examining what happens when chimps are deprived of touch. The chimpanzee was able to touch neither humans, nor its mother, father, or siblings; the animal was destroyed psychologically. A little girl watching the TV show sent Goodall the change purse with a few coins in it, asking her to buy a Snoopy doll for the chimp so it would have something to hug.

Goodall admits that, yes, sometimes she does start to lose hope, but then she pulls out her sacred bundle, and hope returns.

### Evaluation and Sacred Bundles

There’s a line I picked up at an evaluation conference that nonprofits may want to use as funders challenge them to document results: “Measure what you value, and others will value what you measure.” For nonprofits, when it comes to evaluation, the challenge is figuring out what the organization values so dearly that it will track it in real time, so that staff can celebrate what’s working well and make adjustments when things get off track.

We waste a lot of resources when we measure things for the sake of compliance. When we focus, instead, on utilization-focused evaluation, the two critical questions are these: Who’s going to use the information? For what purposes? If nonprofit staff asks only those two questions about evaluation, they’ll do well for themselves.

Nonprofits often focus on the technical aspects of evaluation. What may be missing is the difficult process of reaching consensus on the story an organization wants to tell, that is, on what it values most. In the face of varied and conflicting requirements from funders, it is critical that nonprofits be explicit about their values: That is the only way they will align their outcomes with their values.

1. The story was passed on from Andy Goodman, whose resources on communications are available at [www.agoodmanonline.com](http://www.agoodmanonline.com).

*continued on page 7*

## Assessing Nonprofit Organizational Capacity

*Abby Weiss from HFRP describes the tool that the Marguerite Casey Foundation offers its nonprofit grantees to help them assess their organizational capacity.*

**N**onprofits often are told to maximize their organizational capacity as a way to improve their performance. Knowing what capacity actually means and how to identify areas of need, however, can be confusing. Tools that define and assess organizational capacity can help organizations identify their unique capacity building needs and guide the development of plans to address them.

The Marguerite Casey Foundation—a private foundation that funds community organizing and advocacy with the goal of creating a movement of working families advocating on their own behalf—works with its nonprofit grantees to assess and build their organizational capacity. As part of this process, the foundation offers grantees a self-assessment tool to help them diagnose capacity strengths and challenges, and establish capacity building goals. Results deepen the foundation’s understanding of grantee capacity and help both the foundation and grantees track growth in capacity over time.

The foundation’s Organizational Capacity Assessment Tool is adapted from the McKinsey Capacity Assessment Grid, a comprehensive tool that evaluates organizations on 60 capacity dimensions. Aided by Blueprint Research and Design, Inc., the Marguerite Casey Foundation tailored the grid to fit its own needs and goals, which include encouraging grantees to take stock of their capacity issues, gaining a greater understanding of grantees’ organizational capacity, and informing the foundation’s technical assistance and programming plans.

The modified grid organizes questions under four dimensions: (a) leadership capacity—the ability of the organization’s leaders to inspire, prioritize, make decisions, provide direction, and innovate; (b) adaptive capacity—the ability of the organization to monitor, assess, and respond to changes; (c) management capacity—the ability of the organization’s management to ensure effective and efficient use of resources; and (d) operational capac-

ity—the ability of the organization’s operations to implement key organizational and programmatic functions. The adapted tool includes new questions on community organizing and constituent involvement (which are relevant to the foundation’s movement building work). Blueprint and the Marguerite Casey Foundation also strengthened questions in the areas of evaluation, marketing, communications, fundraising, and cultural competency.

Ideally, several staff members and a board member of the nonprofit grantee each complete the assessment independently, and then meet to compare responses and identify a single set of ratings. This discussion can also include the identification of organizational capacity building goals and action plans. The electronic version of the tool automatically tabulates and summarizes results, facilitating communication among colleagues and allowing organizations to share the data easily with their funders or outside evaluators.

Both the Marguerite Casey Foundation assessment tool and the McKinsey grid are publicly available (see box). Some cautions accompany the tools, which should be treated as “a grading framework” rather than as instruments for precise scientific measurement. First, responses are subjective and respondents can potentially inflate ratings, although rating scales provide for some standardization and encourage candid responses. Second, rating scales can capture only major differences in capacity over time and are not sensitive to smaller increments of progress. Third, because the tools are often customized, data are difficult to aggregate across different versions. However, data can be examined across organizations to determine general areas of strengths and weaknesses.

**Abby R. Weiss**

*Project Manager, HFRP. Email: [abby\\_weiss@harvard.edu](mailto:abby_weiss@harvard.edu)*

### Evaluation and the Sacred Bundle

*continued from page 6*

Three questions may help with this alignment. First, the nonprofit should ask, what story does it want to tell? Organizations should think about that moment down the road when a staff member will have to sit down at the computer and write a few lines about the work. Someone will pull open a drawer, take out a file, and hope that they have the right information.

That brings up the second question: What information does the organization need to tell the story? Organizations must figure out what information they must capture, day by day, so that one day it will be possible to pull that information from a sacred bundle.

The third question is, do organizations currently have in place mechanisms to track that information over time? If so, staff will put the information they value most into their sacred bundle. It will be there when they need it. If the mechanisms are not in place, staff will at least have identified their priority for evaluation.

**John Bare**

*Vice President for Strategic Planning and Evaluation. Arthur M. Blank Family Foundation, 3223 Howell Mill Road, NW, Atlanta, GA 30327. Tel: 404-367-2100. Email: [jbare@ambfo.com](mailto:jbare@ambfo.com)*

### Capacity Building Assessment Resources

**McKinsey Capacity Assessment Grid.** [www.vppartners.org/learning/reports/capacity/assessment.pdf](http://www.vppartners.org/learning/reports/capacity/assessment.pdf)

**Marguerite Casey Foundation Organizational Capacity Assessment Tool.** [www.caseygrants.org/pages/resources/resources\\_downloadassessment.asp](http://www.caseygrants.org/pages/resources/resources_downloadassessment.asp)

Guthrie, K., & Preston, A. (with Hemmings, E., Walker, C., Siqueiros, S., Friedman, E., & Ko, K. L.). (in press). **Building capacity while assessing it: Three foundations’ experiences using the McKinsey Capacity Assessment Grid.** In *A funder’s guide to organizational assessment*. Saint Paul, MN: Fieldstone Alliance. [www.blueprintrd.com/text/capacityassess.pdf](http://www.blueprintrd.com/text/capacityassess.pdf)

## What is strategic learning and how do you develop an organizational culture that encourages it?



*John A. Healy, Director of Strategic Learning and Evaluation at The Atlantic Philanthropies, shares ways to position learning as an organizational priority.*

The Strategic Learning and Evaluation (SLAE) team at The Atlantic Philanthropies develops evaluation techniques, encourages learning from foundation and grantee experiences, and facilitates dissemination of that learning. Strategic learning means that SLAE's audiences—the foundation, grantees, and interested external audiences—understand and respond to the lessons they learn from both formal evaluation and more informal monitoring. We use multiple tactics to create a culture that emphasizes learning.

**1. Position evaluation as a resource linked to strategy.** We began SLAE in early 2004 knowing that it was important to demonstrate the value of learning and evaluation in creating and executing more effective strategy. SLAE staff have both strong evaluation skills and the rare ability to build enthusiasm for evaluation. SLAE works hard to ensure that both grantees and Atlantic see planning and evaluation as strategic and necessary to reach their goals. Within Atlantic we act as a resource, using an approach influenced by strategic planning. SLAE helps staff plan program strategy, map out grant strategies, and ensure programs are ready to be evaluated. We use strategic planning tools (e.g., logic models and theories of change) to clarify our intended outcomes and the strategies to achieve them.

We also work with staff to help grantees plan strategically. Importantly, Atlantic objectives are informed by grantees' objectives, rather than imposed by Atlantic. As a result, grantees have an incentive to learn because their objectives are their own. Once the buy-in for evaluation exists, we position evaluation as practical and useful for grantees' work. For example, human rights grantees' focus has long been on fighting for a particular cause and not on assessment, so we emphasize the importance of "fighting the good fight" *effectively*. If we frame evaluation correctly and align it with the mission of the organization, grantees see it as a win because it informs their work. Our approach helps us adapt as we learn and as the grant making environments change.

**2. Share accountability, acknowledge risk, and reward adaptive learning.** Atlantic has a vertical grant approval structure. Working with grantees, teams of program executives develop grant proposals. They make recommendations to a program investment committee, and the committee submits grants to the board for approval. So at the end of the day, everyone is accountable for decisions about grant making.

Because accountability is shared, it is essential that everyone understands the risks associated with grants and has an open dialogue about what may or may not work. If we are clear on the risks and the grant fails but we learn something from it, the result can be as valuable as if the grant had succeeded. In assessing progress we refer to grantee reports, rigorous project evaluations, and external evaluations of grant clusters in our

four program areas.<sup>1</sup>

Now we are working on creating incentives for learning from mistakes at all levels of the foundation. Staff need to understand that adaptive behavior from learning will be rewarded. Shared accountability helps since the organization is seen as taking the risk rather than individuals.

We also try to be open with grantees about the risks involved in their work. When we encourage evaluation, grantees know we want honest answers. We build relationships with grantees so they see us as advisors, and, if necessary, sources of funding to solve problems. This reinforcing of the message that failure can lead to valuable lessons is essential if a culture of "strategic learning" is to be developed. If the context changes or evaluation findings signal the need for change, grantees understand they can, and should, adjust if their approaches are no longer relevant.

**3. Create space and structure for learning.** It is not enough to encourage learning; we need to create space and structure for doing it. To encourage internal learning, program staff and board members have a fundamental discussion around strategy annually. We ensure the board is clear on what we are trying to achieve, and that they engage in the learning process.

With grantees, we build purposeful opportunities for information exchange. For example, external evaluators meet with Atlantic teams and grantees biannually for a mutual exchange, with evaluators responding to grant-related data that staff and grantees provide and vice versa.

**4. Focus on different levels of learning.** Atlantic has agreed on a number of levels on which we will share learning, and SLAE is now working with our communications staff to ensure learning is targeted effectively to internal and external audiences.

At the foundation level, we will share case studies on organizational development and key areas of change management. Atlantic has chosen to spend down our endowment over the next 10 to 15 years. As we do, we will share lessons on what the foundation does, and what we learn along the way.

Also at the foundation level, we will commission significant studies on areas of strategic importance to the foundation. For example, since many Atlantic objectives focus on policy change, we will share lessons about key aspects of this process, such as evidence gathering and advocacy. Additionally, at the program area level, every 3 to 4 years we will commission more intensive evaluation where we pause and evaluate the core program areas in-depth.

Finally, at the grantee level, we ensure they can share their own lessons. We build in funding for communications, and work closely with grantees to ensure they go that final step.

**Julia Coffman**

*Consultant. HFRP. Email: [jcoffman@evaluationexchange.org](mailto:jcoffman@evaluationexchange.org)*

**Erin Harris**

*Research Analyst. HFRP. Email: [erin\\_harris@harvard.edu](mailto:erin_harris@harvard.edu)*

1. Atlantic's four program areas are Aging, Disadvantaged Children and Youth, Population Health, and Reconciliation and Human Rights.

## What is the Campbell Collaboration and how is it helping to identify “what works”?

*Robert Boruch, a founder of the Campbell Collaboration and professor of education and statistics at the University of Pennsylvania, discusses how the Campbell Collaboration and randomized trials contribute to evidence-based policy.*

The Campbell Collaboration (C2) is a nonprofit organization that aims to help people make well-informed decisions about the effects of interventions in the social, behavioral, and educational arenas. Using systematic reviews of studies of interventions (programs, practices, and policies), C2 helps policymakers, practitioners, researchers, and the public identify what works.

Systematic reviews synthesize available high quality evidence on interventions. After a thorough search of the literature to screen available studies for quality, reviewers identify the least equivocal evidence available on an intervention, describe what the evidence says about the intervention’s effectiveness, and explore how that effectiveness is influenced by variations in process, implementation, intervention components, participants, and other factors.

A main justification for C2’s effort is the volume of studies purporting to show that certain interventions “work.” Often, multiple studies of the same intervention exist, many of which are based on anecdotal or other fragmentary evidence. Reports from such studies can influence decisions about whether to adopt an intervention. Frequently, interventions do not actually have the effects they are purported to have, so decisions to adopt programs may be made based on faulty or incorrect evidence.

C2 reviewers<sup>1</sup> conduct systematic reviews after completing a lengthy protocol that requires that they specify and outline the kinds of outcome variables they will examine, as well as permissible research designs. Virtually all reviews start with the assumption that reviewing randomized trials is the priority: Well-designed randomized trials that are carried out well provide a guarantee that the intervention group (e.g., individuals, schools, classrooms) is similar to the group not exposed to the intervention or exposed to a different one. When resources are available, the C2 reviews also explore high quality nonrandomized trials, but randomized trials are the focus.

An example of a C2 systematic review is one completed on so-called “scared straight programs,” in which kids at risk of committing a crime—and sometimes even those who are not at risk—hear from convicted felons who try to deter them from delinquent acts or crimes. These programs have received a lot of press, are popular with many parents, and they seem to enjoy some approval among policymakers. The systematic review uncovered 200–300 articles from studies of these programs. Only a small fraction turned out to be fair tests of the program. The approximately eight randomized trials identified in the review showed, contrary to popular belief, that these programs actually *enhanced* the likelihood that kids would engage in delinquent behaviors or crimes.<sup>2</sup>

1. Reviewers can come from anywhere in the world and must find their own funding to generate the review.

2. Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2002). “Scared Straight” and

In many governments—not only here in the U.S. but elsewhere as well—the push toward more policy based on less equivocal evidence about the effectiveness of interventions or policies is strong. Often, producing such evidence requires a randomized trial. But does evidence-based policy require the use of randomized trials?

A relatively simple way of outlining the conditions under which evaluators should consider whether a randomized trial is necessary was produced in 1981 by the Federal Judicial Center, the research arm of the U.S. Superior Courts.<sup>3</sup> The Center was interested in whether it was appropriate to do randomized trials in a prison or judicial context. The Center identified five criteria for determining whether a randomized trial is appropriate: (a) the issue must be considered a serious social problem, (b) the solution should be unknown or debatable, (c) no methods other than randomized trials will yield equally defensible evidence, (d) the results should be used in the public sector, and (e) evaluators must be able to protect individual rights involved in the study. If all of these conditions are met, it is sensible to consider doing a randomized trial. If one or more is not met, a randomized trial is probably not appropriate.

**Abby R. Weiss**

*Project Manager, HFRP. Email: [abby\\_weiss@harvard.edu](mailto:abby_weiss@harvard.edu)*

other juvenile awareness programs for preventing juvenile delinquency. *The Cochrane Database of Systematic Reviews* 2002, Issue 2, Art. No.: CD002796.

3. Federal Judicial Center. (1981). *Social experimentation and the law*. Washington, DC: Author.

### Related Resources

Boruch, R. F. (Ed.). (2005). *Place randomized trials: Experimental tests of public policy: Vol. 599. The annals of the American Academy of Political and Social Science*. Thousand Oaks, CA: Sage.

The Campbell Collaborative’s online library offers two databases. **C2 Social, Psychological, Education, and Criminological Trials Registry (C2-SPECTR)** contains over 11,700 entries on randomized and possibly randomized trials. **C2 Reviews of Interventions and Policy Evaluations (C2-RIPE)** contains information on systematic reviews, including titles, protocols, abstracts, and refereed critiques. [www.campbellcollaboration.org/Fralibrary.html](http://www.campbellcollaboration.org/Fralibrary.html)

The **Cochrane Collaboration**, a sibling organization of the Campbell Collaboration, develops systematic reviews in the health care arena. [www.cochrane.org](http://www.cochrane.org)

The **What Works Clearinghouse** offers searchable databases and reports that provide ongoing, high quality reviews of the effectiveness of replicable educational interventions to improve student outcomes. [www.whatworks.ed.gov](http://www.whatworks.ed.gov)

## A conversation with

### Gary Henry

Gary Henry is a professor in the Andrew Young School of Policy Studies and Department of Political Science at Georgia State University. He has evaluated numerous policies and programs, including the Georgia Pre-K and HOPE Scholarship programs, and has published extensively on evaluation methodology and policy analysis, most recently focusing on the effects of education policies, public information campaigns, and, as discussed here, evaluation influence. Henry has been director of evaluation at the David and Lucile Packard Foundation, deputy secretary of education for the Commonwealth of Virginia, and chief methodologist with the Joint Legislative Audit and Review Commission for the Virginia General Assembly. He received the American Evaluation Association Award for Outstanding Evaluation in 1998 and is a former co-editor-in-chief of *New Directions for Evaluation*.

**Q** Evaluators are plagued by the notion that much of our work goes unused. What is your take on how we should think about evaluation use?

**A** Evaluation use has always been important to our field. Most of us consider it an important criterion in judging the merit of our work. We lack clarity, however, in what it means for our work to be used, and consequently the term *use* over time has been stripped of meaning.

Some, for example, define use as direct actions taken as a result of evaluations (also known as instrumental use). Several well-known studies have assessed such “end-state” definitions of use by asking decision makers if they would attribute their actions to evaluations read prior to taking those actions. This research has led to a pervasive notion in the field that evaluations generally go unused.<sup>1</sup> While this definition is one way to think about use, it is too narrow if it is the primary way we think about it. It leaves out other things that evaluation can set in motion.

To take a classic example, over time the 1960s High/Scope evaluation in Ypsilanti, Michigan, has completely changed the way we think about early childhood education.<sup>2</sup> It has become the foundation for the huge movement toward universal preschool that we see now in the United States. We would not acknowledge this kind of use if we just looked at whether Michigan policymakers picked up the evaluation’s results in the year or two after they were released. We would miss this evaluation’s extremely powerful upstream use.

Thinking about use as the end goal of evaluation limits us from

1. E.g., see Weiss, C. H. (Ed.). (1977). *Using social research in public policy making*. Lexington, MA: D.C. Heath.

2. See the High/Scope Educational Research Foundation website for more information on this research ([www.highscope.org](http://www.highscope.org)).

seeing the various types of influence evaluation can have. We need a paradigm shift in how we think about and research this topic. This shift requires that we think about evaluation as an *intervention* with its own set of processes, outputs, and outcomes that we are aware of and accountable for.

**Q** How does thinking about evaluation as an intervention affect how we think about evaluation use?

**A** When thought of as an intervention with its own set of outcomes, it is easier to see that evaluation has the potential to be used in multiple ways. To illustrate this, we can think about mapping the intervention of evaluation like we map other interventions or programs using logic models or theories of change.

First, rather than seeing use as the end goal of evaluation, the end goal should be social betterment. Ultimately, we should be concerned with an evaluation’s influence on the beneficiaries of a program or policy, and look at whether people are better off as a result of the evaluation.

Backing up from this end goal, we can identify the various evaluation outcomes that will lead to social betterment. Rather than view these outcomes as types of evaluation use, however, we should think about them as types of *influence*, or as direct and indirect changes that are triggered by evaluation. By changing the term from use to influence, it becomes easier to think about evaluation as an intervention and to seek the broader ways in which evaluations, or the evaluation process itself, influences social betterment in the long term.

Evaluation outcomes linked to social betterment can be categorized in three ways: individual, interpersonal, and collective. *Individual influence* is when evaluation changes something within

*Thinking about use as the end goal of evaluation limits us from seeing the various types of influence evaluation can have. We need a paradigm shift in how we think about and research this topic.*

the individual, such as one's thoughts, attitudes, beliefs, or actions. *Interpersonal influence* refers to changes triggered by interactions between individuals, such as when an evaluation's findings are used to persuade others about the merit of a program or policy. *Collective influence* means changes in the decisions or practices of organizations or systems, such as when policy change happens as a result of an evaluation, or when a program is expanded, continued, or terminated.

To continue the logic model analogy, we can back up even farther and map the various evaluation activities that can lead to these outcomes. For example, how we select stakeholders, design evaluations, collect and analyze data, generate findings, and disseminate results affects the types of influence our evaluations have.

Reflecting on the model as a whole, we can use it to identify the various pathways through which our evaluations can have influence. For example, in my work on universal preschool a desired social betterment outcome is that kids develop more skills faster as a result of quality preschool programs. To get that outcome, collective action is needed in that public policies have to be in place to support such programs. But other forms of influence also may be needed before support for those policies is generated.

For example, individual influence is important because decision makers must first find the issue compelling and be persuaded about the benefits of universal preschool. Next, decision makers (or other groups) need to become change agents and interpersonally persuade others that preschool is in the best interests of children in their state. From there, the issue has to make it on the policy agenda and be considered by the legislature, administration, and voting public before we eventually get policy reform. If we look only at evaluation in terms of its influence on collective action, however, we miss the other steps and types of influence needed to get there.

**Q** How does thinking about influence change what we do as evaluators?

**A** First, thinking about evaluation influence leads to greater clarity on the purposes of our evaluations. Thinking about influence should encourage us to ask questions about how evaluation sponsors see the evaluation having influence. Is it to inform policy change, and if so, what kind of policy change? Should it inform programmatic change, and, if so, what aspects? As we design evaluations, we should ask these questions and develop at least a mental map of the pathways through which the evaluation can have influence. The evaluation process will be affected by the kinds of influence pursued.

Second, we need to think more about communication strategies. It is important to identify the audiences we need to reach to ensure our work has influence, and to be deliberate actors in achieving that influence. For example, perhaps the issue we're evaluating isn't appearing on the public's or policymakers' radar. We have to consider who we need to reach to raise its visibility, and how our work can affect the issue's salience for individuals in that audience. Or we may need to inspire audiences like the media, advocacy groups, or citizens to make interpersonal use of the evaluation findings. These considerations should be included in evaluation planning, and we should give ourselves enough time to actually follow through on our plans once we make them and have the evaluation findings available.

Third, transparency is important. Underlying the concept of influence is the notion that people have to know evaluation findings in order to use them. An evaluation kept under wraps directly conflicts with this notion. For evaluations completed with both public and private funds, we need to take a careful look at transparency and encourage broad exposure and influence with many audiences.

If evaluators embrace the notion of evaluation as an intervention and are accountable for the outcomes of our work, many types of evaluation influence are possible. Ultimately this can change the way we think about our work and help shed the perception that much of our work goes unused.

**Julia Coffman, Consultant, HFRP**

Email: [jcoffman@evaluationexchange.org](mailto:jcoffman@evaluationexchange.org)



Gary Henry

### Related Resources

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293-314.

Henry, G. T. (2003). Influential evaluations. *American Journal of Evaluation*, 24(4), 515-524.

Henry, G. T. (2000). Why not use? *New Directions for Evaluation*, 88, 85-98.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35-57.



## An Introduction to Theory of Change

Andrea Anderson is a research associate at the Aspen Institute Roundtable on Community Change, where she focuses on work related to planning and evaluating community initiatives.

### What Is a Theory of Change?

A theory of change (TOC) is the product of a series of critical-thinking exercises that provides a comprehensive picture of the early- and intermediate-term changes in a given community that are needed to reach a long-term goal articulated by the community (see box for more detail).

### What Is the Value of Creating a TOC?

Community initiatives are sometimes planned without an explicit understanding of the early and intermediate steps required for long-term changes to occur; therefore, many assumptions about the change process need to be examined for program planning or evaluation planning to be most effective. A TOC creates an honest picture of the steps required to reach a goal. It provides an opportunity for stakeholders to assess what they can influence, what impact they can have, and whether it is realistic to expect to reach their goal with the time and resources they have available.

### What Is a Theory of Change?

A theory of change (TOC) is a tool for developing solutions to complex social problems. A basic TOC explains how a group of early and intermediate accomplishments sets the stage for producing long-range results. A more complete TOC articulates the assumptions about the process through which change will occur and specifies the ways in which all of the required early and intermediate outcomes related to achieving the desired long-term change will be brought about and documented as they occur.<sup>1</sup>

### Steps to Create a Theory of Change

1. Identify a long-term goal.
2. Conduct “backwards mapping” to identify the preconditions necessary to achieve that goal.
3. Identify the interventions that your initiative will perform to create these preconditions.
4. Develop indicators for each precondition that will be used to assess the performance of the interventions.
5. Write a narrative that can be used to summarize the various moving parts in your theory.<sup>2</sup>

1. Adapted from Anderson, A. (2005). *The community builder's approach to theory of change: A practical guide to theory and development*. New York: The Aspen Institute Roundtable on Community Change.

2. Adapted from [www.theoryofchange.org](http://www.theoryofchange.org).

### How Can Initiatives Create a TOC?

Little exists in the way of a methodology for applying the TOC approach to real-world situations. The Aspen Institute Roundtable on Community Change seeks to fill this gap by developing tools and training materials to help program stakeholders create TOCs; we developed a process that community-based initiatives can use to create TOCs.

When I first began to work with groups to create theories of change, most of us thought that TOC was purely an evaluation exercise. But the more I worked with program stakeholders, the more I realized that the exercise of creating a TOC is really an expectation management tool that can benefit planning as well as evaluation; I found that just asking people to articulate assumptions about the change process they hoped to bring about produced a lot of blank stares, or produced work that wasn't quite at the level of critical thinking that we thought was important. So we came up with a step-by-step process that helps to unpack a group's thinking about how to bring about desired changes and how to go about documenting that change in a systematic fashion.

### What Is the Process of Creating a TOC?

The first step is for stakeholders to be clear about what they want to produce through their initiative. We find group members often have very different ideas about what they are working toward. The next step is for stakeholders to think about all of the preconditions—the building blocks or requirements—that must exist in order to reach their long-term goal. They then need to consider, in light of this big picture perspective, which of these preconditions (otherwise known as outcomes) they will take responsibility for producing.

Usually there is just a subset of outcomes that they can influence. Some preconditions are beyond the sphere of influence of any single initiative, such as needing a stable economy to produce enough jobs to reach an employment goal. Others may be beyond one program's influence, but stakeholders could suggest ways that a particular program may be able to influence other programs to act, or they could identify areas for strategic collaboration or partnerships. For example, a precondition for a school readiness initiative might be that all children are properly immunized and healthy before they enter school. A small initiative couldn't influence this precondition, but it may be able to help bring it about through collaboration with others in the community who could directly influence this key precondition for success (see box for a summary of the steps).

### What's the Difference Between a TOC and a Logic Model?

A logic model is a tactical explanation of the process of producing a given outcome. It outlines the program inputs and activities, the outputs they will produce, and the connections between those outputs and the desired outcomes. Alternatively, a TOC, as we define it, is a strategic picture of the multiple interventions required to produce the early and intermediate outcomes that are

*continued on page 19*

## Evaluating Complicated—and Complex— Programs Using Theory of Change

*Patricia Rogers of the Royal Melbourne Institute of Technology describes how a theory of change can provide coherence in evaluating national initiatives that are both complicated and complex.*

Organizations increasingly are adopting and advocating evaluations based on a theory of change (TOC)—an explicit causal model of how program activities are understood to contribute to intended outcomes. It can be difficult, however, to use this approach when planning, managing, and evaluating programs that are both *complicated* and *complex*.

Complicated programs may have multiple parallel or complementary causal paths, that is, alternative ways of achieving outcomes; they may be multilevel programs with local, regional, national, or international layers. Complex programs have at least two main features: emergent properties, where precise objectives and implementation strategies are developed during implementation as specific opportunities develop, and disproportionate relationships, where, at critical levels, a small change can make a big difference, that is, serve as a tipping point.

Capacity-building projects are often both complicated and complex. While the project's overall goals may be clear—for example, building stronger families and communities—the specific activities and causal paths are expected to evolve during implementation; this allows programs to take advantage of emerging opportunities and learn from difficulties.

This type of project calls for a flexible theory of change evaluation. In this type of evaluation, an initial model has been developed and used to guide planning and implementation of a project; however, the theory is revised as plans change. Such flexibility, however, can make it difficult to meet the needs of a national evaluation, which requires some common framework across a set of projects that operate under common goals but are diverse in their local design and implementation. The evaluation of the Australian Government's Stronger Families and Communities Strategy provides an example of how a theory of change can adapt to meet these demands.

### Australia's Stronger Families and Communities Strategy

The Stronger Families and Communities Strategy 2000–2004 (the Strategy)<sup>1</sup> was a national funding program that incorporated over 600 local projects across Australia. Projects were funded through seven funding initiatives, with disbursements ranging from less than \$1,000 to over \$1,000,000 Australian dollars. While all projects were intended to strengthen families and communities, specific project activities varied enormously.

The monitoring and evaluation of the Strategy used a TOC that addressed both the complicated and complex nature of the intervention. The theory of change (see the box) helped provide coherence across the Strategy and could be adapted for evaluating particular projects or clusters of similar projects, or for funding initiatives. This particular TOC (also called an outcomes hier-

archy) delineated seven levels of Strategy outcomes, from short-term to longer term, that capacity building projects could progress through. It was developed as part of a performance indicator framework generated before projects began. Used to develop specific performance indicators, both qualitative and quantitative, for individual projects, the performance indicator framework became the basis of the evaluation framework for the overall Strategy.<sup>2</sup>

**Patricia J. Rogers, Ph.D.**

*Associate Professor in Public Sector Evaluation. Collaborative Institute for Research, Consulting and Learning in Evaluation, Royal Melbourne Institute of Technology, Australia. Email: patricia.rogers@rmit.edu.au*

2. More recently, the evaluation has developed a dynamic theory of change model that explicitly shows the feedback loops in project implementation. The model depicts how project achievements build capacity for further progress, how projects impact family and community strength in the short-term, and, through their contributions to capacity building, their impact in the long-term.

### Stronger Families and Communities Strategy 2000–2004 Theory of Change

The theory of change offers seven levels of Strategy outcomes, from short-term to longer term:

*Level 1: Participation, enhanced trust (via participation, networks, and new ideas) – Includes direct participation in the Strategy, including the application process, even if the application is unsuccessful*

*Level 2: Greater awareness, development of partnerships – Includes awareness of the Strategy, its principles, and values, as well as subject matter awareness to be developed by specific projects*

*Level 3: Greater choice, understanding, skills, capacity for initiative – Includes not just the particular skills, etc., that might have been the direct target of a project but also the understanding, confidence, and capacity acquired by participants in the course of planning and managing projects. Capacity would also include newly established partnerships and so forth. Greater choice could include access to a wider range of services or more appropriate services.*

*Level 4: Demonstration/application of greater understanding, skills, and capacity – Includes an element of sustainability beyond the completion of the funding period*

*Level 5: Family and community trust/resilience/adaptability – About trust that would transcend the particular project (whereas Level 1 might be about trust developed on a smaller scale through a particular Strategy project)*

*Level 6: Environment where communities participate in and drive their own solutions to strengthen themselves – Participation at this level transcends participation that occurs in relation to a particular project*

*Level 7: Stronger families and communities – About how families and communities apply the strengths from Levels 1 through 6 to improve their well-being*

1. For more information on the evaluation of the Strategy see [www.facs.gov.au/internet/facsinternet.nsf/aboutfacs/programs/sfsc-evaluation.htm](http://www.facs.gov.au/internet/facsinternet.nsf/aboutfacs/programs/sfsc-evaluation.htm).

## The Knight Foundation's Approach to Cluster Evaluation

*The John S. and James L. Knight Foundation and Wellsys Corporation describe how they plan to aggregate lessons learned across a “thematic cluster” of youth development investments.*

In an effort to support adolescents during the critical middle school transition years, the John S. and James L. Knight Foundation, through its Community Partners Program, made significant investments in out-of-school-time programming in 6 of the 26 communities in which it works.<sup>1</sup> At each site, Knight's local Community Advisory Committees (CACs)—groups of community, business, and nonprofit leaders who help define funding priorities that fit the foundation's mission and strategically address issues of community concern—have focused the foundation's grantmaking on distinct but related efforts to improve outcomes for at-risk adolescents.

### The Six Communities in the Cluster

The six communities vary considerably on a number of dimensions—size, location, demographics, and nonprofit capacity—and have found unique ways to target their grantmaking, often focusing on specific schools, neighborhoods, or ZIP codes. Additionally, although the six committees all based their funding decisions on similar concerns about the challenges facing young people, each approached grantmaking differently. In Milledgeville, Georgia, Knight is supporting the replication of a nationally recognized youth development and pregnancy prevention program. In Columbus, Georgia, Knight is helping existing youth-serving organizations develop new, collaborative models for service delivery. In Palm Beach, Florida, the foundation is focused on building the capacity of the youth development field and ensuring that community-based after school providers achieve the level of quality necessary for achieving positive youth outcomes.

With local funding strategies in place, Knight staff began investigating ways to aggregate lessons learned across this thematic cluster of investments. The foundation contracted with the Atlanta-based Wellsys Corporation, whose staff recently began working to develop a baseline view of the grant cluster as well as an ongoing evaluation plan.<sup>2</sup>

### Developing the Baseline View

Wellsys, through its initial review of foundation grantmaking documents, has found several different approaches and strategies to help paint the baseline picture. Within each community with a youth development focus, Wellsys encountered many components, such as grant foci, community reach, theory of change, participants, service delivery systems, and embedded evaluations that needed to be unpacked and documented in order to really be able to assess the complex cluster. Additional components should emerge as conversations take place with foundation staff, local grantees, and evaluators. The cluster evaluation plan will rely on

the local evaluators and established evaluation plans that many of the communities already have in place—a purposeful attempt to utilize local resources to the fullest possible extent.

Wellsys is also using several other strategies to develop the youth development cluster baseline view. The first was to interview Knight's Community Liaison Program Officers to document and understand progress to date and help determine next steps for engaging the identified communities. During the summer of 2005, the Wellsys team will visit each community to meet key grantees and their corresponding evaluators. Through these site visits, the team hopes to learn about the evaluation scope and methods being used in each setting; gain a deeper understanding of grantee activities; and obtain input on the cluster evaluation process. Grantee buy-in during the evaluation's early stages will be crucial, as the foundation hopes that grantees will benefit from the assessment through cross-community learning exchanges that may take the form, for example, of technical assistance and recommendations for improving programs and evaluation systems.

Following these community visits, Wellsys will develop a preliminary document that describes early evaluation findings and current status of evaluations, and outlines what the foundation may realistically glean about grantee effectiveness and impact at the end of the 2-year study period. In November 2005 this report will be shared with Knight Foundation's staff and board of trustees, as well as local evaluators and representatives from each community, to obtain feedback. Knight Foundation hopes that this cluster, or “systems,” view will lead to greater insight into the strengths and challenges of its various funding strategies.

Once the cluster theory of change has been developed, Wellsys, in concert with Knight Foundation, will identify the process and outcome evaluation questions and conduct an analysis to note any gaps in community data collection. Tools will be developed to capture identified information and processes established to support data collection. This aggregation of data will inform and support a higher level of foundation decision making and provide additional insight into the utility of various youth development programs in producing clearly identified outcomes and leveraging broader community change.

**Julie K. Kohler, Ph.D.**

*Interim Director, National Venture Fund. Email: kohler@knightfdn.org*

**Lizabeth Sklaroff**

*President's Special Assistant for Evaluation.  
Email: sklaroff@knightfdn.org*

*The John S. and James L. Knight Foundation, Wachovia Financial Center, Suite 3300, 200 South Biscayne Boulevard, Miami, FL 33131-2349*

**Denise Townsend, Ph.D.**

*President and Co-Founder. Email: dtownsend@wellsyscorp.com*

**Susan Boland Butts**

*Senior Consultant. Email: sbbutts@bellsouth.net*

*Wellsys Corporation, 5045 Johns Creek Court, Alpharetta, GA 30022*

1. The six communities are Akron, Ohio; Bradenton and Palm Beach County, Florida; Columbia, South Carolina; and Columbus and Milledgeville, Georgia.

2. Both the baseline view and evaluation plan are slated for completion by December 2005, with the evaluation plan to be implemented over the next 2 years.

## Ten Strategies for Enhancing Multicultural Competency in Evaluation

*Teresa Boyd Cowles of the Connecticut Department of Education offers self-reflective strategies evaluators can use to enhance their multicultural competency.*

Many evaluators—from the novice to the experienced—will find themselves involved with evaluating programs for a growing number of diverse and multicultural children, youth, families, and communities. Growth in diversity demands multiculturally competent evaluators, yet evaluators' efforts to enhance their skills in working with multicultural populations are not keeping up with the demand.

Cultural competence has been described as “a set of congruent behaviors, attitudes, and policies that come together in a system, agency, or among professionals and enables that system, agency, or those professionals to work effectively in cross-cultural situations.”<sup>1</sup> The prefix *multi* and the suffix *culture* imply many integrated patterns of human behavior, including thoughts, communications, actions, customs, beliefs, and values as they relate to racial, ethnic, religious, or social groups. *Competence* implies the capacity to function in a particular way within the context of culturally integrated patterns of human behavior. Consequently, competent multicultural functioning means learning new patterns of behavior and effectively applying them in appropriate or varied settings.

Continuous self-reflection and self-evaluation are proactive means for sustaining multicultural competency. The following strategies have been adapted from Gorski's suggestions for becoming a better multicultural educator and serve as a guide for enhancing evaluators' multicultural competency to improve evaluation practice.<sup>2</sup>

**1. Engage in self-reflective writing or journaling to explore your own identity development.** Everyone has a multicultural identity. Reflect on how your identity has developed and evolved over time. Erikson began writing about identity development years ago<sup>3</sup> and many others since have agreed that an authentic racial identity is a critical component to becoming culturally competent.<sup>4</sup>

1. Cross, T. L., Bazron, B. J., Dennis, K. W., & Isaacs, M. R. (1989). *Towards a culturally competent system of care: A monograph on effective services for minority children who are severely emotionally disturbed* (Vol. 1). Washington, DC: Georgetown University Center for Child and Human Development.

2. Gorski, P. (2004). 20 (self-)critical things I will do to be a better multicultural educator. Retrieved June 8, 2005, from [http://www.edchange.org/multicultural/resources/self\\_critique.html](http://www.edchange.org/multicultural/resources/self_critique.html)

3. Erikson, E. H. (1968). *Identity: Youth and crisis*. London: Faber & Faber.

4. Carter, R. T. (1995). *The influence of race and racial identity in psychotherapy: Toward a racially inclusive model*. San Francisco: Jossey Bass; Giroux, H. A. (1997). Rewriting the discourse of racial identity: Towards a pedagogy and politics of Whiteness. *Harvard Educational Review*, 67(2), 285–320; Helms, J. E. (Ed.). (1990). *Black and white racial identity: Theory, research, and practice*. Westport, CT: Greenwood; Howard, G. R. (1999). *We can't teach what we don't know: White teachers, multiracial schools*. New York: Teachers College Press; Phinney, J. (1992). The Multigroup Ethnic Identity Measure (MEIM): A new scale for use with adolescents and young adults from diverse groups. *Journal of Adolescent Research*, 7(2), 156–176; Tatum, B. D. (1994). Teaching white students about racism: The search for White allies and the restoration of hope. *Teachers College Record*, 95(4), 462–476.

**2. Invite critique from colleagues and accept it openly.** It's difficult to accept critique and even more difficult to publicly solicit and acknowledge feedback. Get in the habit of inviting critique and thanking colleagues for offering it. Sharing with your peers enables you to be an objective, fair, and judicious evaluator.

**3. Examine the relationship between intent and impact.** We intend evaluations to be beneficial; however, their impact can be detrimental. It is easier to focus on our intent regardless of the impact, but unintentional outcomes can be just as damaging as intentional ones. Realize that there may be a mismatch between your intent and the actual impact of an evaluation, and reflect on how to take a different approach in the future.

**4. Reject the myth of color blindness.** It is simply unnatural to be color blind. Everyone is touched, moved, or affected by race the moment they enter the world. Race shapes how others see us and how we view ourselves. Achievement gaps, dropout rates, and life expectancy are a reality of being a person of color in the United States. Seeing color acknowledges the racial disparities and affirms the whole person; color is an important part of our identity.

**5. Recognize that stakeholders' identity group may affect their experiences and the evaluation process.** Evaluation may be an unwelcome part of any community or organization, and evaluators may have little control over how people experience or interpret their intentions. Stakeholders base their reactions on a lifetime of experiences (i.e., their own identity development). Try not to take stakeholders' reactions personally, but be aware that they may have an impact on the evaluation process. As stated by Scriven, “The most difficult problems with program evaluation are not methodological or political but psychological.”<sup>5</sup>

**6. Build coalitions with evaluators who differ in race, ethnicity, sexual orientation, gender, religion, first language, disability, and other identity facets.** Most of us grew up or live in relatively homogenous communities and may only interact with people who are different from us at work or at professional meetings. Go beyond superficial socializing by consciously developing multicultural coalitions, even in informal settings. This effort will foster authentic, trusting, and honest relationships.

**7. Acknowledge the role of a social activist.** Evaluators help determine the merit, worth, and effectiveness of programs and policies that affect multicultural populations. Because of this, evaluators can change the lives of others. Acknowledging this privately and publicly is an important step in accounting for the responsibility, power, and privilege that evaluators have.

**8. Educate yourself.** Multicultural competence is about what is yet to be discovered in the moment; it is contextually specific. Recognize the limitations of textbook cultural information. Hav-

5. Scriven, M. (Ed.). (1993). Hard-won lessons in program evaluation [Special issue]. *New Directions for Evaluation*, 58.

*continued on page 19*

## The Evidence Base for Increasing High-Achieving Minority Undergraduates

*Mehmet Öztürk discusses findings from a review of evaluations of programs at selective colleges and universities to be used for improving undergraduate academic outcomes for underrepresented minority or disadvantaged students.*

African American, Latino, and Native American undergraduates are severely underrepresented in the student bodies of the nation's selective colleges and universities.<sup>1</sup> They also are acutely underrepresented among graduates with high grade point averages and in terms of class rankings.<sup>2</sup> These patterns can continue into graduate and professional schools.<sup>3</sup>

The Consortium for High Academic Performance (CHAP) at the Institute for the Study of Social Change at the University of California, Berkeley, spent 2 years gathering information on programs designed to improve academic outcomes for underrepresented minority or disadvantaged undergraduates. Over 100 strategies and programs housed at public and private research universities and private liberal arts colleges were examined nationwide. Most, though not all, programs focused on science, math, or engineering majors. Methods included extensive literature and document review, site visits, and dialogue with program directors, faculty, senior administrators, and students.

The purpose was to identify *proven* or *promising* programs that could be replicated. Proven programs were defined as those demonstrating measurable impacts on grade point average (GPA), according to an evaluation using an experimental design that had to have been replicated in at least one additional site. Promising programs were defined as those that did not meet these design criteria but showed promising GPA outcomes.

The search uncovered no proven programs; it did, however, uncover several promising programs with sufficient GPA data to suggest that they contributed to meaningful increases in the number of above-average to top performers.<sup>4</sup>

The scarcity of rigorous evaluations for these types of programs can be traced to several possible causes. One is shortage of resources. Most programs operate on very tight budgets and have small staffs; thus, resources tend to be invested directly in the students being served. Resource constraints also limit the development of new or modified programs. Few external funders (foundations and government agencies) invest systematically in ongoing development and evaluation of these programs, particularly from a high-achievement (i.e., high GPA) perspective. Furthermore, few selective colleges and universities seem to invest much of their own resources in operating or evaluating programs serving these

students. Operating programs with quality over time is therefore difficult.

Similarly, institutional mechanisms for training professionals to operate programs demonstrating (or showing promise of) high-achievement impacts are lacking. Thus, even if proven programs were to emerge and efforts made to use them widely, developing and maintaining staff could remain problematic.

These and other findings are described in detail in a recent CHAP report, *Increasing African American, Latino, and Native American Representation Among High Achieving Undergraduates at Selective Colleges and Universities*.<sup>5</sup> The report emphasizes the need for leadership from presidents of selective colleges and universities, heads of foundations with an interest in improving the academic outcomes of underrepresented students in higher education, and leaders of organizations concerned with the educational advancement of underrepresented minorities.

According to the report, leaders will need to focus on establishing large, long-term grant programs to support the development and evaluation of programs that address high achievement among these student populations. Leaders also will need to secure stable funding (e.g., endowments) to sustain proven and promising programs. Finally, they will need to establish mechanisms for training professionals to replicate programs.

Additionally, the report illustrates the kinds of rigorous evaluations needed by offering specific design suggestions using three promising programs as examples. Also provided are examples of variations of these programs that need to be tested and evaluated.

### Mehmet D. Öztürk, Ph.D.

*Principal Management Research Analyst, Office of the Vice President for University-School Partnerships, Office of the Dean, College of Education, Arizona State University. Tel: 480-965-3696. Email: ozturk@asu.edu*

5. Miller L. S. (with Öztürk, M. D., & Chavez, L.). (2005). *Increasing African American, Latino, and Native American representation among high achieving undergraduates at selective colleges and universities*. Berkeley: University of California, Berkeley, Institute for the Study of Social Change.

1. U.S. News & World Report. (2005). *America's best colleges: 2005 edition*.

2. Cole, S., & Barber, E. (2003). *Increasing faculty diversity: The occupational choices of high-achieving minority students*. Cambridge, MA: Harvard University Press.

3. See the following for more information: Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press; and Sander, R. H. (2005). A systematic evaluation of affirmative action in American law schools. *Stanford Law Review*, 57(3), 376-483.

4. Other promising programs may exist that were not identified for this search.

### Upcoming Issues of The Evaluation Exchange

Democratic Evaluation – Fall 2005

Professional Development – Winter 2005/2006

Linkages in Out-of-School Time – Spring 2006

To subscribe go to [www.gse.harvard.edu/hfrp/subscribe.html](http://www.gse.harvard.edu/hfrp/subscribe.html).

To submit information email [hfrp\\_pubs@gse.harvard.edu](mailto:hfrp_pubs@gse.harvard.edu).

### Recent Issues of The Evaluation Exchange

Complementary Learning – Spring 2005

Family Involvement – Winter 2004/2005

Harnessing Technology – Fall 2004

Read or print these issues at [www.gse.harvard.edu/hfrp/eval/archives.html](http://www.gse.harvard.edu/hfrp/eval/archives.html).

## Building a Pipeline Program for Evaluators of Color

*Rodney Hopson and Prisca Collins of Duquesne University describe a new graduate internship program designed to develop leaders in the evaluation field and improve evaluators' capacity to work responsively in diverse racial and ethnic communities.*

In 1999 the American Evaluation Association (AEA) launched the Building Diversity Initiative (BDI),<sup>1</sup> which produced a number of recommendations both for the AEA and the evaluation field as a whole. One important outcome of the 2-year initiative was the recommendation and subsequent establishment of the Graduate Education Diversity Internship Program at Duquesne University.

### Program Background

Advisory committee and initiative staff surveyed evaluators<sup>2</sup> and discovered a lack of mentoring opportunities, role models, and access to training for evaluators of color. In addition, they interviewed foundation and federal agency representatives about their engagement with evaluators of color. These interviews exposed the difficulties these institutions face in ensuring culturally responsive evaluations. Respondents attributed this, in part, to the challenge of identifying, accessing, and engaging both evaluators of color and evaluators with the capacity to work with racially and ethnically diverse communities. In fact the majority of the respondents' institutions were not engaged in deliberate efforts to identify diverse evaluators.

Increasing the number of evaluators of color is critical not only because of the new ideas, paradigms, and realities these professionals provide but also because of the significance of cultural context and competence in the field. In recent years, there has been a growing movement to apply cultural litmus tests to evaluation processes, standards, and use, particularly when communities of color are participants and stakeholders.<sup>3</sup>

One factor contributing to the low number of evaluators of color is the lack of available training opportunities. Despite increased demands for accountability and evaluation in foundation and government sectors, the number of formal graduate programs in evaluation has been decreasing over the last decade, with pre-

dictions that these programs are unlikely to expand.<sup>4</sup> Moreover, there is a dearth of doctoral degrees earned by African Americans, American Indians, Mexican Americans, and Puerto Ricans in research-based education and the social sciences.<sup>5</sup>

There are several possible approaches to building innovative training mechanisms for increasing diversity among evaluators. Frierson<sup>6</sup> suggests that schools of education play a major role in both educating and training program evaluators and increasing the diversity of the field. Discussions from a recent National Science Foundation workshop on the role of minority evaluation professionals suggest expanding the current model to a multiple-agency approach; such a structure might include the involvement of colleges and universities, government agencies, and professional organizations through mentoring, potential employment, and other strategies.<sup>7</sup>

### Duquesne University's Graduate Internship Program

As a result of the BDI, AEA and Duquesne University established a graduate internship program that is intended to (a) build the "pipeline" of students who already have basic research capacities and substantive knowledge and extend their capabilities to evaluation, (b) provide professional development training opportunities for social science, public health, and other graduate students in research, and (c) deepen the evaluation profession's capacity to work in racially, ethnically, and culturally diverse settings.

In its inaugural year (2004–2005), the program was guided by a project director, internship coordinator, and advisory board comprising leaders from AEA and the larger evaluation field. The program's accomplishments to date include the following:<sup>8</sup>

- Recruiting and enrolling graduate students of color
- Developing the internship curriculum, Evaluation Toward Social Justice and Social Change, which guided three training sessions facilitated by internship staff and evaluation experts
- Placing interns with local agencies where they participated in evaluating programs serving culturally diverse populations
- Advising and providing interns with mentor support from an

*Increasing the number of evaluators of color is critical not only because of the new ideas, paradigms, and realities these professionals provide but also because of the significance of cultural context and competence in the field.*

1. The initiative was staffed by the Association for the Study and Development of Community (ASDC).

2. Of the 1,047 respondents, 50% were European American, 19% African American, 10% Latino, and 8% Asian American.

3. Some of the literature that has contributed to this movement includes the following: Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. *New Directions for Evaluation*, 92, 31–43; Hopson, R. K. (2003). *Overview of multicultural and culturally competent program evaluation: Issues, challenge, and opportunities*. Woodland Hills, CA: The California Endowment; Thompson-Robinson, M., Hopson, R., & SenGupta, S. (Eds.). (2004). In search of cultural competence in evaluation [Special issue]. *New Directions for Evaluation*, 102.

4. Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. Boston: Allyn & Bacon.

5. Frierson, H. T. (2003). The importance of increasing the numbers of individuals of color to enhance cultural responsiveness in program evaluation. In C. C. Yeakey & R. D. Henderson (Eds.), *Surmounting all odds: Education, opportunity, and society in the new millennium*. Greenwich, CT: Information Age; National Science Foundation. (2000). *The cultural context of educational evaluation: The role of minority evaluation professionals*. Arlington, VA: Author. [www.nsf.gov/pubs/2001/nsf0143/start.htm](http://www.nsf.gov/pubs/2001/nsf0143/start.htm)

6. Frierson, 2003.

7. National Science Foundation, 2000.

8. The program's first year was made possible with invaluable support from the W. K. Kellogg Foundation, the Duquesne University School of Education, and the American Evaluation Association.

*continued on page 19*

## The Retrospective Pretest: An Imperfect but Useful Tool

*Theodore Lamb, of the Center for Research and Evaluation at Biological Sciences Curriculum Study, discusses retrospective pretests and their strengths and weaknesses.*

Evaluators often use pretests and posttests in their evaluation designs, taking measures before and after an intervention to see what difference the intervention made. At times, however, evaluators can find such designs difficult to implement, especially in interventions such as professional development (PD) workshops. In such workshops it can be burdensome to administer instruments twice, or worse, the traditional pretest may not be effective if participants do not sufficiently understand, prior to the workshop, terms or concepts needed to answer pretest questions. Campbell and Stanley<sup>1</sup> offer a solution in the form of the *retrospective pretest* (RPT).

The RPT is administered at the same time as the posttest, that is, respondents are asked to answer questions about their level of understanding or skill *after* an intervention, such as a PD workshop, occurs. They are then asked to think back to their understanding *prior* to the intervention; this is the retrospective pretest. Selected strengths and weaknesses of using the RPT method are discussed below (see also box).

### Retrospective Pretest Strengths

*Single administration.* Workshop participants sometimes dislike completing evaluation forms before and again after a workshop. One clear advantage of the RPT is that evaluation forms are administered only once, at the end of the workshop. This single administration saves time and may be more satisfying to participants.

*Avoiding the response-shift effect.* The RPT can also help to avoid a response-shift effect,<sup>2</sup> which occurs when a respondent's frame of reference or evaluation standard changes significantly during an intervention. Consider, for example, a workshop for teachers focused on the topic of inquiry-based teaching. If workshop participants misunderstand basic terms or concepts associated with inquiry-based teaching, then results from traditional pretest questions may be misleading.

The RPT method avoids the response-shift effect by clearing up misconceptions before participants are asked to make assessments. Once a workshop is over and all concepts have been sufficiently explained, participants first assess their new level of understanding or skill, and, second, reflectively assess the level of understanding or skill they had prior to the workshop.

### Retrospective Pretest Weaknesses

*Desire by participants to show a learning effect.* A "good subject effect"<sup>3</sup> occurs when subjects try to figure out what a researcher

wants and then give responses to support what they believe that to be. The RPT may introduce a permutation of the good subject effect: the *learning effect*, or the tendency for workshop participants to show, in order to help the workshop deliverers look good, that learning took place regardless of whether or not it did.

*Threats to validity.* Self-report data and the recall of information through reflection may be subject to problems of insufficient recall as well as offer the potential for fabricated or biased responses.

The logic and value of experimental designs has a long and distinguished history in many disciplines and should be used when they can help to answer research or evaluation questions. However, if for some reason a traditional pretest and posttest cannot be used, then the RPT is better than either not evaluating at all or taking only posttest measures.

To explore possible differences between traditional pretests/posttests and the RPT, we collected data in several PD workshops using both types of administrations. Results showed little differences between the two approaches. We concluded therefore that the RPT is a good method to use if it is difficult or impossible to use traditional pretests. However, we recommend that the RPT be supplemented by additional data gathered to illustrate the effectiveness of workshops and other interventions.

### Theodore Lamb

*Co-Director, Center for Research and Evaluation, Biological Sciences Curriculum Study, 5415 Mark Dabling Blvd., Colorado Springs, CO 80918. Tel: 719-531-5550. Email: tlamb@bscs.org*

### Strengths and Weaknesses of Retrospective Pretests

#### Strengths

- Less likely to offend participants who do not like being put in the role of students or research subjects required to complete both pretests and posttests
- Can be used when traditional pretests are not possible
- Unlike traditional pretests, does not risk negatively impacting intervention effectiveness by possibly introducing terms and concepts before participants are ready for them
- Provides data that, with other supporting data, can be used to evaluate the effectiveness of a professional development intervention

#### Weaknesses

- May introduce a desire for participants to show a learning effect
- Challenges traditional methodological logic, since both predata and postdata are collected after the intervention
- May introduce threats to validity such as memory recall, history, and regression to the mean
- Possibility of fabricated and biased responses
- Can be perceived as less rigorous, and therefore less convincing, than other approaches

1. Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

2. Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self reports. *Evaluation Review*, 4(1), 93–106.

3. Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783.

## Theory of Change

*continued from page 12*

preconditions of reaching an ultimate goal.

Once a precondition (or outcome) has been identified through the TOC process, a logic model can be used to explain how that outcome will be produced. Thus, one TOC could actually be linked to a number of logic models, because a logic model could be constructed to illustrate how to produce each outcome in the TOC map. The TOC summarizes work at a strategic level, while a logic model would be used to illustrate the tactical, or program-level, understanding of the change process. A Microsoft PowerPoint presentation on the TOC website, [www.theoryofchange.org](http://www.theoryofchange.org), explains this issue in more detail.

### How Are Concepts of TOCs and Logic Models Evolving?

During the early to mid-1990s, funders began increasing their emphasis on outcomes and accountability. As a result, people began paying more attention to TOCs and logic models and often created their own definitions of these concepts to meet their needs. Because people are increasingly seeing the value of doing this kind of work and because it's not a "branded" idea, there are more definitions of these concepts now than there were 10 years ago. My concern is not that the field decide to do it one way, but that people use the exercise of creating a theory of change to ask hard questions about why they expect certain interventions to bring about the outcomes they seek, to question their assumptions about how the change process will unfold, and to be clear about how they're selecting outcomes to focus on.

Instead of thinking about TOC as the "magic bullet" that can solve all of their planning and evaluation challenges, I hope that people will use TOC to learn how to ask fundamentally different—and much more interesting—questions from those we are used to asking about any initiative. The advice I would give to somebody coming at TOC for the first time is to be open-minded about the extent to which it can help them to be better strategic thinkers throughout all of their work, and not to think of it as *just* a planning or evaluation tool.

*TOC tools and training materials are available on the Theory of Change website, a joint venture between the Roundtable and Act-Knowledge: [www.theoryofchange.org](http://www.theoryofchange.org).*

#### Erin Harris

*Research Analyst, HFRP. Email: [erin\\_harris@harvard.edu](mailto:erin_harris@harvard.edu)*

---

## Enhancing Multicultural Competency

*continued from page 15*

ing adept knowledge about the history and current struggles of the social and cultural groups that compose society is fundamental to being affirmative and inclusive. However, education also comes from stepping out of your comfort zone. Explore different people on multiple dimensions and in different places; this is education.

**9. Take responsibility before looking for fault elsewhere.** Stakeholders who appear to be purposely uncooperative may be simply

reacting to your body language, management style, history from past evaluations, or other factors you bring to the table. Consider what may or may not be contributing to their disengagement. Taking accountability for your role in your interactions during the evaluation process improves the conditions of the evaluation.

**10. Examine your methods.** Your rationale for selecting particular methods and tools serves to enhance or inhibit what Kirkhart terms *multicultural validity*.<sup>6</sup> Scrutinizing your methodology will make you more familiar with a variety of techniques, thereby enhancing your ability to distinguish, detect, and control measurement bias in testing and other assessment tools.

"Culture is dynamic and ever-changing; therefore enhancing multicultural competency is a life-long process."<sup>7</sup> Unless evaluators engage in critical self-reflection and ongoing discovery, they stay trapped in unexamined judgments, interpretations, assumptions, and expectations. Every moment spent in self-critique, however challenging and tedious, gives rise to an evaluator who is better equipped to work with the multicultural people and diverse communities that encompass evaluation practice. Welcome to a never ending journey.

#### Teresa Boyd Cowles, Ph.D.

*Education Consultant. Program Evaluation, Connecticut State Department of Education, Connecticut Technical High School System, 25 Industrial Park Road, Middletown, CT 06457. Tel: 860-807-2224. Email: [teresa.boyd@po.state.ct.us](mailto:teresa.boyd@po.state.ct.us)*

6. Kirkhart, K. E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice*, 16(1), 1–12.

7. Symonette, H. (2004). Walking pathways toward becoming a culturally competent evaluator: Boundaries, borderlands, and border crossings. *New Directions for Evaluation*, 102, 95–109.

---

## Evaluators of Color

*continued from page 17*

advisor at students' home institutions and an active senior AEA member with similar research or career interests

- Providing leadership development opportunities in AEA and beyond through participation at the annual AEA meeting, exposure to former and current AEA leaders through trainings and seminars, and active participation in other evaluation-related activities

*Eligibility and application procedures for the internship program are available at [www.education.duq.edu/mastersCertification/diversity\\_inter\\_program.htm](http://www.education.duq.edu/mastersCertification/diversity_inter_program.htm).*

#### Rodney Hopson

*Associate Professor and Project Director. Tel: 412-396-4034. Email: [hopson@duq.edu](mailto:hopson@duq.edu)*

#### Prisca Collins

*Internship Coordinator. Tel: 412-396-5568. Email: [moetip@duq.edu](mailto:moetip@duq.edu)*

*Duquesne University, Department of Educational Foundations and Leadership, School of Education, Canevin Hall, Pittsburgh, PA 15208-0540*

*This section features an annotated list of papers, organizations, initiatives, and other resources related to the issue's theme.*

Callahan, S. (2005). *Singing our praises: Case studies in the art of evaluation*. Washington, DC: The Association of Performing Arts Presenters. This book highlights examples of how arts presenters have used participatory evaluation to learn about their success. Real-life stories, guides, and techniques from other fields are used to train practitioners to design their own evaluations. [www.fortharts.org/book.shtml](http://www.fortharts.org/book.shtml)

The **Evaluation Checklists Website** from the Evaluation Center at Western Michigan University is a compilation of checklists for conducting various aspects of an evaluation, with the goal of improving evaluation quality and consistency and enhancing evaluation capacity. Written by evaluation experts, the checklists are brief distillations of lessons learned from practice. Checklist categories include evaluation management, models, values and criteria, and meta-evaluation. At least 30 checklists are available, and the site is regularly updated. [www.wmich.edu/evalctr/checklists](http://www.wmich.edu/evalctr/checklists)

The **Interactive Smart Chart** is a free online planning tool from Spitfire Strategies that helps nonprofits develop high-impact communications strategies. It allows users to take stock of their responses and progress as they move through the planning process, and to stop and save answers at any time. By the end users have a complete communications strategy—referred to as a Smart Chart—that links overall organizational goals to the many strategic decisions necessary for a successful communications effort. The online tool also builds evaluation into the planning process, and, along with the communications strategy, at the end prints a logic model that maps measures of effort, effect,

and impact to the strategy. A print version of the Smart Chart is also available. [www.smartchart.org](http://www.smartchart.org)

The *Journal of Multidisciplinary Evaluation* is a new peer-reviewed online journal published in association with the doctoral program in evaluation at The Evaluation Center at Western Michigan University. This journal aims to communicate about evaluation to a very diverse readership, and therefore focuses on much more than just theory or empirical research. The journal explores new niches by pursuing significant evaluation happenings from countries outside of North America, translating articles into different languages, and accepting both traditional and nontraditional article formats. [evaluation.wmich.edu/jmde](http://evaluation.wmich.edu/jmde)

Funded by the National Science Foundation, the **Online Evaluation Resource Library's Professional Development Modules** offers self-guided training on a variety of evaluation topics, including evaluation design, methodological approaches and sampling, questionnaires, and interviews. Designed both for novice and experienced evaluators, the modules present step-by-step strategies with scenarios and case studies of how strategies can be applied to specific evaluation projects. [oerl.sri.com](http://oerl.sri.com)

Werner, A. (2004). *A guide to implementation research*. Washington, DC: Urban Institute Press. This toolkit for program evaluation provides a framework for organizing implementation studies and for understanding specific research strategies. The guide also addresses the range of implementation research questions and techniques available to program evaluators and the results that may be expected. Examples of implementation research on welfare programs are included. [www.uipress.org](http://www.uipress.org)



Harvard Family Research Project  
Harvard Graduate School of Education  
3 Garden Street  
Cambridge, Massachusetts 02138  
[hfrp@gse.harvard.edu](mailto:hfrp@gse.harvard.edu)  
[www.hfrp.org](http://www.hfrp.org)

nonprofit org.  
US postage  
**PAID**  
Boston, MA  
Permit #56376

