

Celebrating Our 10th Year!



the

# evaluation exchange

A PERIODICAL ON EMERGING STRATEGIES IN EVALUATING CHILD AND FAMILY SERVICES

HARVARD FAMILY RESEARCH PROJECT ■ HARVARD GRADUATE SCHOOL OF EDUCATION ■ VOL. IX NO. 4 ■ WINTER 2003/2004

## from the director's desk



Heather Weiss

With this issue, *The Evaluation Exchange* kicks off its tenth year of publication. In our very first issue, I said that the success of this resource would depend on your willingness to be an active participant. A decade later, I want to thank you for heeding that call in ways that have exceeded our expectations.

Thank you to the hundreds of authors who have generously

shared their experiences and thoughts, to the thousands of subscribers who have read and applied the content, and to the generous funders that have supported the production and free distribution of *The Evaluation Exchange*. Finally, I want to thank my staff for remaining committed to *The Evaluation Exchange* and to growing its value.

Rarely does a nonprofit publish a periodical that sustains beyond its inaugural issues. We are extremely proud that *The Evaluation Exchange* has evolved into a nationally known resource, with a growing and diverse audience of more than 13,000 evaluators, practitioners, policy-makers, and funders.

While we want to celebrate this tenth-year milestone, we know that we are operating in a dynamic environment with ever-changing demands and appetites for new ideas presented in new ways. We can't stop to celebrate too long—we need to be reflecting constantly on our field and practice so we can continuously improve our work.

Accordingly, we have dedicated this issue to sharing some of the lessons that will inform our agenda in the future. We begin in our Theory & Practice section with a series of reflections by renowned experts on what the past decade has meant for evaluation. These essays point to areas where we have not come far enough, identify

practices we may need to rethink and, in addressing the importance of learning from our progress and success, introduce a theme that emerges several times in this issue.

Several articles offer additional thoughts about recent developments. Michael Scriven offers his perspective on the status of the evaluation profession and discipline. Other articles present nominations for the "best of the worst" evaluation practices, emerging links between program evaluation and organization development, and some surprising findings about changes in university-based evaluation training.

Building on these and our own reflections, this issue also introduces topics that future issues will address in more depth. While our basic format and approach will remain the same, we have included articles that herald our commitment to covering themes we think require more attention in the evaluation arena—diversity, international evaluation, technology, and evaluation of the arts. Upcoming issues will feature and spur dialogue about these topics and others, including program theory, mixed methods, democratic approaches to evaluation, advocacy and activism, accountability, and systems change, which was the topic of our first issue 10 years ago and remains a significant evaluation challenge today.

Like the evaluation field itself, much has changed about *The Evaluation Exchange*. But our success still depends on the participation of our readers. If you have ideas about other topics you would like to see featured, please don't hesitate to share them with us. We continue to welcome your feedback and contributions.

## in this issue

### *Reflecting on the Past and Future of Evaluation*

#### **Theory & Practice**

Experts recap the last decade in evaluation 2

#### **Ask the Expert**

Michael Scriven on evaluation as a discipline 7

#### **Promising Practices**

Evaluators' worst practices 8

Efforts to link evaluators worldwide 12

Narrative methods and organization development 14

#### **Questions & Answers**

A conversation with Ricardo Millett 10

#### **Beyond Basic Training**

Evaluation training trends 13

#### **Book Review**

*The Success Case Method* 16

#### **Spotlight**

Evaluation to inform learning technology policy 17

#### **Evaluations to Watch**

Evaluating a performing and visual arts intervention 18

#### **New Resources From HFRP** 19

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



**Founder & Director**

Heather B. Weiss, Ed.D.

**Managing Editor**

Julia Coffman

**HFRP Contributors**

Heather B. Weiss, Ed.D.

Julia Coffman

Tezeta Tulloch

**Publications/Communications  
Manager**

Stacey Miller

**Publications Assistant**

Tezeta Tulloch

©2004 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,  
W. K. Kellogg Foundation, and  
John S. and James L. Knight  
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/  
submission.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.

**Where We've Been and Where We're Going:  
Experts Reflect and Look Ahead**

*In this special edition of Theory & Practice, six evaluation experts share their thoughts on how the field has progressed (or regressed) in the last 10 years and consider what the next steps should be.*

Articles that appear in Theory & Practice occupy an important position and mission in *The Evaluation Exchange*. Tied directly to each issue's theme, they lead off the issue and provide a forum for the introduction of compelling new ideas in evaluation with an eye toward their practical application. Articles identify trends and define topics that deserve closer scrutiny or warrant wider dissemination, and inspire evaluators and practitioners to work on their conceptual and methodological refinement.

An examination of the topics covered in Theory & Practice over the last decade reads like a chronicle of many of the major trends and challenges the evaluation profession has grappled with and advanced within that same timeframe—systems change, the rise of results-based accountability in the mid-1990s, advances in mixed methods, learning organizations, the proliferation of complex initiatives, the challenges of evaluating communications and policy change efforts, and the democratization of practices in evaluation methodology, among many others.

As this issue kicks off the tenth year of publication for *The Evaluation Exchange*, Harvard Family Research Project is devoting this section to a discussion of some of the trends—good and bad—that have impacted the theory and practice of evaluation over the last 10 years.

We asked six experts to reflect on their areas of expertise in evaluation and respond to two questions: (1) Looking through the lens of your unique expertise in evaluation, how is evaluation different today from what it was 10 years ago? and (2) In light of your response, how should evaluators or evaluation adapt to be better prepared for the future?

**On Theory-Based Evaluation:  
Winning Friends and Influencing People**

Carol Hirschon Weiss

*Professor, Harvard University Graduate School of Education, Cambridge, Massachusetts*

One of the amazing things that has happened to evaluation is that it has pervaded the program world. Just about every organization that funds, runs, or develops programs now calls for evaluation. This is true locally, nationally, and internationally; it is almost as true of foundations and voluntary organizations as it is of government agencies. The press for evaluation apparently arises from the current demand for accountability. Programs are increasingly called on to justify their existence, their expenditure of funds, and their achievement of objectives. Behind the calls for accountability is an awareness of the gap between almost unlimited social need and limited resources.

An exciting development in the last few years has been the emergence of evaluations based explicitly on the theory underlying the program. For some time there have been exhortations to base evaluations on program theory. I wrote a section in my 1972 textbook urging evaluators to use the program's assumptions as the framework for evaluation.<sup>1</sup> A number of other people have written about program theory as well, including Chen,<sup>2</sup> Rossi (with Chen),<sup>3</sup> Bickman,<sup>4</sup> and Lipsey.<sup>5</sup>

<sup>1</sup> Weiss, C. H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.

<sup>2</sup> Chen, H. T. (1990). Issues in constructing program theory. *New Directions for Program Evaluation*, 47, 7-18.

For a while, nothing seemed to happen; evaluators went about their business in accustomed ways—or in new ways that had nothing to do with program theory. In 1997 I wrote an article, “How Can Theory-Based Evaluation Make Greater Headway?”<sup>6</sup> Now theory-based evaluation seems to have blossomed forth. A number of empirical studies have recently been published with the words “theory-based” or “theory-driven” in the titles (e.g., Crew and Anderson,<sup>7</sup> Donaldson and Gooler<sup>8</sup>). The evaluators hold the program up against the explicit claims and tacit assumptions that provide its rationale.

One of the current issues is that evaluators do not agree on what “theory” means. To some it refers to a down-to-earth version of social science theory. To others, it is the logical sequence of steps necessary for the program to get from here to there—say, the steps from the introduction of a new reading curriculum to improved reading. Or, when a violence prevention program is introduced into middle school, it’s what has to happen for students to reduce the extent of bullying and fighting. To others it is the plan for program activities from start to completion of program objectives, without much attention to intervening participant actions.

I tend to see “theory” as the logical series of steps that lays out the path from inputs to participant responses to further intervention to further participant responses and so on, until the goal is achieved (or breaks down along the way). But other evaluators have different conceptualizations. It is not necessary that we all agree, but it would be good to have more consensus on what the “theory” in theory-based evaluation consists of.

Program theory has become more popular, I think, because, first, it provides a logical framework for planning data collection. If a program is accomplishing what it intends to at the early stages, it is worth following it further. If the early phases are not realized (e.g., if residents in a low-income community do not attend the meetings that community developers call to mobilize their energies for school improvement), then evaluators can give early feedback to the program about the shortfall. They need not wait to collect data on intermediate and long-term outcomes if the whole process has already broken down.

<sup>3</sup> Chen, H. T., & Rossi, P. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7(3), 283–302.

<sup>4</sup> Bickman, L. (Ed.). (1987). Using program theory in evaluation. *New Directions for Program Evaluation*, 33, 5–18.

<sup>5</sup> Lipsey, M. W. (1993). Theory as method: Small theories of treatments. *New Directions for Program Evaluation*, 57, 5–38.

<sup>6</sup> Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, 21(4), 501–524.

<sup>7</sup> Crew, R. E., Jr., & Anderson, M. R. (2003). Accountability and performance in charter schools in Florida: A theory-based evaluation. *American Journal of Evaluation*, 24(2), 189–212.

<sup>8</sup> Donaldson, S. I., & Gooler, L. E. (2003). Theory-driven evaluation in action: Lessons from a \$20 million statewide work and health initiative. *Evaluation and Program Planning*, 26(4), 355–366.

A second reason has to do with complex programs where randomized assignment is impossible. In these cases, evaluators want some way to try to attribute causality. They want to be able to say, “The program caused these outcomes.” Without randomized assignment, causal statements are suspect. But if evaluators can show that the program moved along its expected sequence of steps, and that participants responded in expected ways at each step of the process, then they can claim a reasonable approximation of causal explanation.

A third advantage of theory-based evaluation is that it helps the evaluator tell *why* and *how* the program works. The evaluator can follow each phase posited by the theory and tell which steps actually connect to positive outcomes and which ones are wishful fancies.

I intended to end with an inspiring call to evaluators to test out theory-based evaluation more widely and share their experiences with all of us. When we understand the problems that beset efforts at theory-based evaluation, perhaps we can improve our understanding and techniques. But I’ve run out of space.

### On Methodology: Rip Van Evaluation and the Great Paradigm War

Saumitra SenGupta  
Research Psychologist,  
Department of Public Health,  
City & County of San Francisco

Rip Van Evaluation fell asleep in 1991 right after the American Evaluation Association (AEA) conference in Chicago. The qualitative-quantitative battle of the Great

Paradigm War was raging all around him. Lee Sechrest, a supporter of the quantitative approach, had just delivered his presidential speech at the conference<sup>9</sup> in response to the one given by qualitative advocate Yvonna Lincoln<sup>10</sup> the year before.

Fast-forward a dozen years to 2003. At the AEA conference in Reno, the battle lines had become blurred and evaluators were no longer picking sides. David Fetterman was presenting on empowerment evaluation at the business meeting of the Theory-Driven Evaluation topical interest group,<sup>11</sup> interspersed among presentations by Huey-Tsyh Chen,<sup>12</sup> Stewart Donaldson,<sup>13</sup> Mel Mark,<sup>14</sup> and John Gargani.<sup>15</sup> Rip was awakened by

<sup>9</sup> Sechrest, L. (1992). Roots: Back to our first generations. *Evaluation Practice*, 13(1), 1–7.

<sup>10</sup> Lincoln, Y. (1991). The arts and sciences of program evaluation. *Evaluation Practice*, 12(1), 1–7.

<sup>11</sup> Fetterman, D. M. (2003). *Theory-driven evaluation with an empowerment perspective*. Paper presented at the American Evaluation Association annual conference, Reno, NV.

<sup>12</sup> Chen, H. T. (2003) *Taking the perspective one step further: Providing a taxonomy for theory-driven evaluators*. Paper presented at the American Evaluation Association annual conference, Reno, NV.

<sup>13</sup> Donaldson, S. I. (2003). *The current status of theory-driven evaluation: How it works and where it is headed in the future*. Paper presented at the American Evaluation Association annual conference, Reno, NV.

*The past decade can be  
rightfully characterized as  
one where qualitative  
methods, and consequently  
mixed-method designs, came  
of age, coinciding with and  
contributing to a different  
understanding of what  
evaluative endeavor means.*

Saumitra SenGupta,  
Department of Public Health,  
San Francisco

the loud applause while Jennifer Greene was receiving AEA's esteemed Lazarsfeld award for contributions to evaluation theory and mixed methods. Rip's shock at what he was witnessing caps the last 10 years in evaluation.

The past decade can be rightfully characterized as one where qualitative methods, and consequently mixed-method designs, came of age, coinciding with and contributing to a different understanding of what evaluative endeavor means. Understanding, recognizing, and appreciating the context and dimensions of such endeavors have become more salient and explicit. The "value-addedness" of qualitative methods has consequently become more apparent to evaluators, which in turn has made mixed-method designs commonplace—the currency of the day.

Rossi notes that the root of this debate lies in the 1960s<sup>16</sup> and, while the "long-standing antagonism"<sup>17</sup> was somewhat suppressed with the formation of the AEA, it became more prominent during the early 1990s through the Lincoln-Sechrest debate, characterized as "the wrestlers" by Datta.<sup>18</sup> Reichardt and Rallis credit David Cordray, AEA President in 1992, for initiating a process for synthesis and reconciliation of the two traditions.<sup>19</sup>

House cautions the evaluator against becoming fixated with methods and the accompanying paradigm war. While acknowledging the importance of methods, House argues for giving the content of the evaluative endeavor the limelight it deserves.<sup>20</sup> Datta provides a more historical perspective, arguing that the contrast between qualitative and quantitative methods was not as dichotomous as it was being made out to be.<sup>21</sup> Nor was it accurate to portray a theorist-specific espousal of a particular methodology, as most of these theorists and practitioners have from time to time advocated for and employed other methods.

The successful integration started taking root with the pragmatists proposing incorporation of both types of methods for the purposes of triangulation, expansion, and validation, among others.<sup>22</sup> These efforts were reinforced at a conceptual

<sup>14</sup> Mark, M. M. (2003). Discussant's remarks presented at the Theory-Driven Evaluation Topical Interest Group business meeting at the American Evaluation Association annual conference, Reno, NV.

<sup>15</sup> Gargani, J. (2003). *The history of theory-based evaluation: 1909 to 2003*. Paper presented at the American Evaluation Association annual conference, Reno, NV.

<sup>16</sup> Rossi, P. (1994). The war between the quals and the quants: Is a lasting peace possible? *New Directions for Program Evaluation*, 61, 23–36.

<sup>17</sup> Reichardt, C. S. & Rallis, S. F. (1994). Editors' notes. *New Directions for Program Evaluation*, 61, 1.

<sup>18</sup> Datta, L. E. (1994). Paradigm wars: A basis for peaceful coexistence and beyond. *New Directions for Program Evaluation*, 61, 53–70.

<sup>19</sup> Reichardt & Rallis 1.

<sup>20</sup> House, E. R. (1994). Integrating the quantitative and qualitative. *New Directions for Program Evaluation*, 61, 13–22.

<sup>21</sup> Datta 53–70.

<sup>22</sup> Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(2), 13–17; Greene, J. C., Caracelli, B. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274.

level by Cordray<sup>23</sup> and were truly integrated by methods such as concept mapping whereby a fundamentally quantitative method is used in conjunction with focus groups and interpretive methods.<sup>24</sup>

The fruit of this early sowing can be seen in what Rip Van Evaluation woke up to in Reno. There have been many notable developments and advancements in evaluation theory and practice in areas such as ethics, cultural competence, use of evaluation, and theory-driven evaluation. The relationship between organizational development and evaluation theory has also been better defined, through areas such as learning organizations. But the acceptance of mixed-method designs and qualitative methods as coequal and appropriate partners has been a giant step.

### On Evaluation Use: Evaluative Thinking and Process Use

Michael Quinn Patton  
Faculty, Union Institute and University,  
Minneapolis, Minnesota

A major development in evaluation in the last decade has been the emergence of *process use* as an important evaluative contribution.<sup>25</sup> Process use is distinguished from findings use and is indicated by changes in

thinking and behavior, and program or organizational changes in procedures and culture stemming from the learning that occurs during the evaluation process. Evidence of process use is represented by the following kind of statement after an evaluation: "The impact on our program came not just from the findings, but also from going through the thinking process that the evaluation required."

This means an evaluation can have dual tracks of impact: (1) use of findings and (2) helping people in programs learn to think and engage each other evaluatively.

Teaching evaluative thinking can leave a more enduring impact from an evaluation than use of specific findings. Specific findings typically have a small window of relevance. In contrast, learning to think evaluatively can have an ongoing impact. Those stakeholders actively involved in an evaluation develop an increased capacity to interpret evidence, draw conclusions, and make judgments.

Process use can contribute to the quality of dialogue in community and program settings as well as to deliberations in the national policy arena. It is not enough to have trustworthy and accurate information (the informed part of the informed citizenry). People must also know how to use information, that is, to weigh evidence, consider contradictions and inconsistencies, articulate values, and examine assumptions, to note but a few of

<sup>23</sup> Cordray, D. S. (1993). Synthesizing evidence and practices. *Evaluation Practice*, 14(1), 1–8.

<sup>24</sup> Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, 12(1), 1–16.

<sup>25</sup> Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.



the things meant by *thinking evaluatively*.

Philosopher Hannah Arendt was especially attuned to critical thinking as the foundation of democracy. Having experienced and escaped Hitler's totalitarianism, she devoted much of her life to studying how totalitarianism is built on and sustained by deceit and thought control. In order to resist efforts by the powerful to deceive and control thinking, Arendt believed that people needed to *practice thinking*.

Toward that end she developed eight exercises in political thought. Her exercises do not contain prescriptions on what to think, but rather on the critical processes of thinking. She thought it important to help people think conceptually, to "discover the real origins of original concepts in order to distill from them anew their original spirit which has so sadly evaporated from the very keywords of political language—such as freedom and justice, authority and reason, responsibility and virtue, power and glory—leaving behind empty shells."<sup>26</sup> We might add to her conceptual agenda for examination and public dialogue such terms as performance indicators and best practices, among many evaluation jargon possibilities.

From this point of view, might we also consider every evaluation an opportunity for those involved to practice thinking? Every utilization-focused evaluation, by actively involving intended users in the process, would teach people how to think critically, thereby offering an opportunity to strengthen democracy locally and nationally.

This approach opens up new training opportunities for the evaluation profession. Most training is focused on training evaluators, that is, on the supply side of our profession. But we also need to train evaluation users, to build up the demand side, as well as broaden the general public capacity to think evaluatively.

### On Evaluation Utilization: From Studies to Streams

Ray C. Rist

*Senior Evaluation Officer, The World Bank, Washington, D.C.*

For nearly three decades, evaluators have debated the variety of uses for evaluation. An evaluation has been generally understood to be a self-contained intellectual or practical product intended to answer the information needs of an intended user. The unit of analysis for much of this navel gazing has been the single evaluation, performed by either an internal or external evaluator and presumably used by stakeholders in expanding concentric circles. The debate about the use—and abuse—of evaluations has thus hinged on what evidence can be mustered to support evaluations' direct, instrumental "impact" or "enlightenment." Evaluators have attempted to identify other forms of use as well, such as conceptual/illuminative, persuasive,

<sup>26</sup> Arendt, H. (1968). *Between past and future: Eight exercises in political thought* (pp. 14–15). New York: The Viking Press.

and interactional. These debates have reflected a notion of evaluations as discrete studies producing discrete "findings."

The most recent developments have focused on new notions of process use or, notably, "influence." Still, this debate among evaluators on the use of evaluations has been essentially a closed one, with evaluators talking only among themselves. Sadly, those involved seem oblivious to fundamental changes in the intellectual landscape of public management, organizational theory, information technology, and knowledge management.

But this era of endless debates on evaluation utilization should now end. New realities ask for, indeed demand, a different conceptualization about evaluation utilization.

We are in a new period where ever-accelerating political and organizational demands are reframing our thinking about the definition of what, fundamentally, constitutes evaluation and what we understand as its applications. This period is characterized by at least two important considerations. The first is the emergence of an increasingly global set of pressures for govern-

ments to perform effectively—to go beyond concerns with efficiency—and to demonstrate that their performance is producing desired results. The second is the spread of information technology, which allows enormous quantities of information to be stored, sorted, analyzed, and made available at little or no cost. The result is that where governments, civil societies, and policymakers are concerned, the value of individual evaluations is rapidly diminishing.

The issue is no longer the lack of evidence for instrumental use by those in positions of power who could make a difference. Rather, users of evaluative knowledge are now confronted with growing rivers of information and analysis systematically collected through carefully built monitoring systems. Users are fed with streams of information from the public, private, and nonprofit sectors in country after country across the globe. Witness the following four examples:

1. The budget system in Chile (and soon in Mexico as well), which links evaluation performance information to budget allocations on a yearly basis.
2. The 24-hour monitoring system in New York City on policing patterns and their relation to crime prevention and reduction.
3. Databases in the United States continuously updated on the performance of medical devices.
4. Continuous assessment of different poverty alleviation strategies in developing countries (Uganda being notable in this regard).

These examples suggest the range of evaluative information systems currently being built and deployed. None depends on individual, discrete evaluation studies.

Increasingly, a large proportion of contemporary organizations and institutions thrive on their participation in the knowledge processing cycle. (By focusing on the multiple dimensions

*These organizations now talk not about [data] scarcity, but about managing the information deluge. Use becomes a matter of applying greater and greater selectivity to great rivers of information.*

Ray C. Rist,  
*The World Bank*

of the knowledge processing cycle, they are seeking to bypass the endless discussions on what is knowledge and what is information. They want to stay out of that cul-de-sac.) These organizations and institutions understand and define themselves as knowledge-based organizations, whatever else they may do, be it sell insurance, teach medical students, fight AIDS, or build cell phones. In fact, and somewhat ironically, these organizations now talk not about scarcity, but about managing the information deluge. Use becomes a matter of applying greater and greater selectivity to great rivers of information.

Far from concentrating on producing more and more individual evaluation studies, we see that governments, nongovernmental organizations, and the private sector are all using new means of generating real-time, continuous flows of evaluative knowledge for management and corporate decisions. These new realities completely blur and make obsolete the distinctions between direct and indirect use, between instrumental and enlightenment use, and between short and long term use.

*The views expressed here are those of the author and no endorsement by the World Bank Group is intended or should be inferred.*

### **On Community-Based Evaluation: Two Trends**

Gerri Spilka  
*Co-Director, OMG Center for Collaborative Learning,  
Philadelphia, Pennsylvania*

With over 10 years of community-based evaluation experience under our belts at OMG, I look back at a range of evaluations—from formative to impact, and from ones focused on local and area-wide programs to broader national initiatives and “cluster evaluations” that reviewed entire areas of grantmaking intended to change a particular system in a region. Examples of these include evaluations of the Comprehensive Community Revitalization Program (CCRP) initiative in New York’s South Bronx, the Annie E. Casey Foundation’s Rebuilding Communities Initiative (RCI), the Fannie Mae Foundation’s Sustained Excellence Awards Program, and, more recently, a citywide cluster evaluation of grantmaking to support the development of various community responsive databases. As I look back, two big trends stand out. One represents a big shift; the other, a shift that has not gone far enough.

The big change has to do with who is listening to the conversations and the evaluation reports about community-based work. When OMG first completed the CCRP evaluation in 1993, we had a limited audience that included only a select group of other evaluators, grantmakers, and community-based activists. But around that time the knowledge dissemination potential of the Internet was rapidly becoming apparent, a change that helped support the expanding use of professional networks. Professionals in the field were developing a huge appetite for new practical knowledge of effective strategies and the Internet now provided the means to share it easily. As evaluators we became credible sources of opinion about effective community programs and in many cases we found ourselves

brokering information as a new, valued commodity.

Also during this time, for a number of reasons, policy-makers started listening to us more. They read our reports and checked our sites; we had their ear. Eager to advance change in their own communities, they wanted evidence of successful programs to turn them into policy. It became even more critical for us to demonstrate benefits in real outcomes—real numbers and real dollars saved.

Another trend that has appeared over the past decade, but that has thus far not borne enough fruit, is the increasing attention to outcomes thinking throughout the field. The problem here is that despite the new outcomes fascination, progress has been slow in harnessing this thinking to improve practice. Particularly troubling is our own and our clients’ inability to be realistic about what kinds of outcomes we can expect from the work of the best minds and hearts of community activists within the timeframes of most grants and programs.

Ten years ago, five million dollars in a community over eight years seemed like a lot of money and a long commitment. We hoped we would see incredible outcomes as the result of these investments. Our first-generation logic models for comprehensive community revitalization efforts included, among many others, changes such as “reductions in drug-related crime, existence of an effective resident governance of public human services, and increases in employment.” Our good intentions, sense of mission, and optimism set us up to expect dramatic neighborhood change in spite of decades of public neglect. Nor did we always realistically factor in other community variables at play. In many cases, we also expected inexperienced and under-capacitated community-based organizations to collect data for us—an assumption that, not surprisingly, led to great disappointment in the quality of the data collected.

Sadly, despite lots of experience to draw from, we have not yet developed a thorough understanding of what constitute reasonable outcomes for these initiatives, nor have we come to agree on the most effective ways to collect the data essential to sound evaluation. As a result, we still run the risk of making poor cases for the hard and passionate work of those struggling to improve communities. Gary Walker and Jean Baldwin Grossman recently captured this dilemma well. They argue that the outcomes movement has come of age and that never before have foundations and government been so focused on accountability and outcomes. Accountability and learning about what works is a good thing. However, “even successful...programs rarely live up to all the expectations placed in them.”<sup>27</sup>

As we look to the future, being realistic about outcomes and measuring them effectively remain challenges. In the near-term political environment it may prove harder to make our case. But we do now have the ears of policymakers in an unprecedented way. We have the means to influence more people about what is possible with the resources available. We must be rigorous, not just about measuring results, but also about setting expectations for what is possible.

*continued on page 20*

<sup>27</sup> Walker, G., & Grossman, J. B. (1999). *Philanthropy and outcomes: Dilemmas in the quest for accountability*. Philadelphia: Public Private Ventures.



## Michael Scriven on the Differences Between Evaluation and Social Science Research

*Michael Scriven is a professor of evaluation at the University of Auckland in New Zealand and a professor of psychology at Claremont Graduate University in California. One of the world's most renowned evaluators, Dr. Scriven has authored more than 330 publications in 11 fields, including Evaluation Thesaurus, a staple of evaluation literature. Dr. Scriven is a former president of the American Evaluation Association (AEA), and received the AEA's esteemed Lazarsfeld Award for his contributions to evaluation theory.*

### How are evaluation and social science research different?

Evaluation determines the merit, worth, or value of things. The evaluation process identifies relevant values or standards that apply to what is being evaluated, performs empirical investigation using techniques from the social sciences, and then integrates conclusions with the standards into an overall evaluation or set of evaluations (Scriven, 1991).

Social science research, by contrast, does not aim for or achieve evaluative conclusions. It is restricted to empirical (rather than evaluative) research, and bases its conclusions only on factual results—that is, observed, measured, or calculated data. Social science research does not establish standards or values and then integrate them with factual results to reach evaluative conclusions. In fact, the dominant social science doctrine for many decades prided itself on being value free. So for the moment, social science research excludes evaluation.<sup>1</sup>

However, in deference to social science research, it must be stressed again that without using social science methods, little evaluation can be done. One cannot say, however, that evaluation is the application of social science methods to solve social problems. It is much more than that.

### What unique skills do evaluators need?

Evaluators need a few special empirical research skills along with a range of evaluative skills. The repertoire of empirical skills mainly includes those used for social science research, with its emphasis on hypothesis testing. But for an evaluator, empirical skills must include more than those required for traditional social science research.

For example, evaluators often need to know how to search for a program or policy's side effects—a skill that is tremendously important for evaluation, but not for hypothesis testing. For an evaluator, discovering side effects may be what swings the overall evaluative conclusions from bad to good or vice versa.

Evaluative skills also include abilities like determining relevant technical, legal, and scientific values that bear on what is

<sup>1</sup> Note, however, that this is changing as social science is being asked to be more involved with serious social problems, interventions, or issues. In order to do so, social science will have to incorporate evaluation or evaluative elements.

being evaluated, and dealing with controversial values and issues.

Evaluators also need synthesis skills in order to integrate relevant evaluative and factual conclusions. In fact, the ability to synthesize is probably the key cognitive skill needed for evaluation. Synthesis includes everything from making sure that judgments are balanced to reconciling multiple evaluations (which may be contradictory) of the same program, policy, or product (Scriven, 1991).

### Why aren't the differences between evaluation and social science research widely understood or accepted?

One has to understand the difference between a profession and a discipline. Program evaluation began to take shape as a profession during the 1960s and has become increasingly “professional” in the decades since. This progress has mostly involved the development of evaluation tools, the improved application of these tools, the growth of a professional support network, and a clearer understanding of the evaluator's status and role. This is very different from what it takes to develop into a discipline.

A discipline recognizes the boundaries of a field and its relation to other fields. It has a concept of itself, as well as an appropriate philosophy of operation that defines the logic of that particular field. The recognition that allows a profession to be thought of as a discipline comes well after that profession has developed. For evaluation that recognition has come only recently.

Evaluation's move toward becoming a discipline was delayed by the prominence of the value-free doctrine in the standard social sciences centering on the assertion that evaluation could not be objective or scientific and therefore had no place as a scientific discipline. It was not until the late twentieth century that this thinking was confronted seriously and its inaccuracies discovered.

While evaluation has been practiced for many years, it is only now developing into a discipline. In this way evaluation resembles technology, which existed for thousands of years before there was any substantive discussion of its nature, its logic, its fundamental differences from science, and the details of its distinctive methods and thought.

In recent years we have begun to see more discussions within the field about evaluation-specific methodology. We are moving toward the general acceptance of evaluation as a discipline, but there is still a long way to go.

### Reference

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

**Julia Coffman, Consultant, HFRP**  
Email: [julia\\_coffman@msn.com](mailto:julia_coffman@msn.com)

## Looking the Enemy in the Eye: Gazing Into the Mirror of Evaluation Practice

*David Chavis, President of the Association for the Study and Development of Community, outlines the “best of the worst” evaluator practices when it comes to building relationships with evaluation consumers.<sup>1</sup>*

Being an evaluator is not easy. I’m not referring to the technical problems we face in our work, but to how people react to us and why. Telling someone that you’re an evaluator is like telling them you’re a cross between a proctologist and an IRS auditor. The news evokes a combination of fear, loathing, and disgust, mixed with the pity reserved for people who go where others don’t want them to go.

I have developed this perspective through providing hundreds of evaluators and evaluation consumers with technical assistance and evaluation training, through overseeing the “clean up” of evaluations by “prestigious” individuals and institutions, and through conducting many evaluations myself.

In each case I heard stories about those evaluators that make evaluation consumers look at all of us with contempt, suspicion, and, on good days, as a necessary evil. I began to consider, who is this small minority messing things up for the rest of us?

Just about every evaluator I spoke with said it was the previous evaluator that made his or her work so difficult. I realized that, to be occurring at this scale, these bad experiences either grew out of an urban myth or were the work of a band of renegade, number crunching, ultra-experimentalist, egomaniac academics.

Then it hit me—it’s all of us. Maybe we all are contributing to this problem. As Laura Leviton said in her 2001 presidential address to the American Evaluation Association (AEA), “Evaluators are not nice people.” As nice as we might be privately, we generally don’t know how to build and maintain mutually enhancing and trustful relations in our work. Threads on EvalTalk, the AEA listserv, frequently demonstrate the difficulties we have getting along.

Many evaluators think consumers react negatively to us out of fear we will reveal that their programs aren’t working as well as they think, and because they believe we have some special access to The Truth. Think about how you’d feel if someone talked to you for five minutes and then told you how to improve yourself. Who has the ability, or the nerve, to do that?

I’ve seen evaluators borrow from telephone psychics to help consumers overcome these fears—we provide insights no one can disagree with, like: “Your funders don’t understand what

you are trying to accomplish,” and follow with the clincher: “I am really excited by your work.”

But are we? Or are we excited about what we can gain from their work? Are we fellow travelers on the road to the truth about how to improve society or are we just about the wonders of the toolbox (i.e., methods)?

Bruce Sievers and Tom Layton<sup>2</sup> recognize that while we focus on best practices we neglect worst practices, even though we can learn a lot from them—especially about building better relations. In the interest of learning, the following are some of the worst evaluator and participant relationships my colleagues and I have seen.

### Not Listening or Not Acting on What We’ve Heard

Stakeholders often tell us that although they want something useful from their evaluation—not something that “just sits on a shelf”—all they get is a verbal or written report. When challenged we say they should have paid for it or given us more time. We also often hear that practitioners are afraid negative results will affect their funding. We assure them we’ll take care of it, while thinking to ourselves there’s nothing we can do. After all, we aren’t responsible for negative results—we just tell the truth.

In most of these cases the evaluator simply hasn’t listened and thought through how to deal with the situation. Often the

stakeholders’ real question is: Will you struggle with us to make this program better, or will you just get the report done?

It is essential to conduct active and reflective interviews with stakeholders. We need to agree on how we can improve their program as part of the evaluation process. Even if there is a small budget, that relationship-building time must be considered as important as data analysis.

### Branding: Evaluation As a Package Deal

In a world where the label on your jeans says who you are, it’s not surprising evaluators sell “brands” of evaluation. At the recent AEA meeting, six leaders in the field presented their brands. They recognized some overlap, but emphasized the uniqueness of their approaches. For me, each approach reflected a different decision I might make on any given day, but I couldn’t see a big difference. What I fear is having to describe to evaluation consumers what I do based on these brands: “I’m doing a theory-driven, responsive, transformative, utilization-focused, empowerment, goal-free, collaborative, participatory, outcome-focused evaluation.” Would I still have anybody’s attention?

*Being an evaluator is not easy ... Telling someone that you’re an evaluator is like telling them you’re a cross between a proctologist and an IRS auditor.*

<sup>1</sup> Any similarities to individual evaluators are unfortunate, but coincidental. I make gross generalizations that apply to all of us, though admittedly to some more than others.

<sup>2</sup> Sievers, B., & Layton, T. (2000). Best of the worst practices. *Foundation News and Commentary*, 41(2), 31–37.



Often the brands represent rigid packages that become more important than their appropriateness within a specific context. In one recent situation, when the program operators told the evaluator that his method wasn't working for their organization, he responded that they didn't understand his approach and shouldn't have agreed to it and that, after all, it's working just fine all over the world—just read his book.

As evaluators, we must get beyond our brands and labels, except for one: multi-method. The needs of the context must be the driving force, not the promotion of our latest insight or evaluation package.

### Keeping Our Methods a Mystery to Maintain Our Mastery

David Bakan,<sup>3</sup> a noted philosopher of science, described the mystery-mastery effect as one that psychologists use to maintain their power over people. When we present our methods, analysis, and our ability to be objective in a way that's above the understanding of the public, we create power over the consumer. Did we take a course in keeping our opinions to ourselves and making "objective" judgments? No, but we would hate to dispel the myth of our objectivity—without it, what would make us special? We need to dedicate ourselves to educating the public on the diversity of our methods, our approach to knowledge development, and the limitations and profound subjectivity of our work.

### Thinking We Can Hold Everyone Accountable But Ourselves

Many evaluators think we should be allowed to do what we see fit—that we need neither monitoring nor review of our work. Many consumers think we do exactly what we want to do. As evaluators, we are getting what we want, although it may not be well liked by others. Many other professionals have systems of accountability, including physicians, accountants, lawyers, and architects. Even if these systems are flawed, their mere existence shows that these professionals and the public hold their work in high esteem.

Contracting problems are plentiful in the evaluation field. Evaluators still frequently enter relations without contracts specifying the deliverables. There are widespread misunderstandings over the end results of the evaluator's work. How do we hold ourselves accountable? Is it just driven by the market? (I.e., as long as someone is paying you, you're cool?) We need to evaluate our own work in the same manner we profess to be essential for others.

### Going It Alone While Overlooking Our Limitations

I have my own pet theories. All professions can be divided up into dog professions and cat professions. Law, for example, is a

dog profession—lawyers work well in packs and they bark. Evaluation is a cat profession—independent, aloof, sucks up to no one. Plus, we evaluators know how to get out quick and hide when there's a loud noise.

There is great pressure on us to know and do everything. We are asked to facilitate, conduct strategic planning sessions and workshops, produce public information documents, and give advice, frequently without much training or direct experience ourselves. Rarely do I see us working in teams, let alone with other "experts." We tend to go it alone, giving it the ol' educated guess. We need to develop relations with other experts with complementary practices.

### Forgetting We Are Part of the Societal Change Process

The work we evaluate exists because there are people out there who have a deep passion to change society, or their little piece of it. Often we see those with passion as more biased, more motivated by self-interest, and less knowledgeable than ourselves. When practitioners criticize the sole use of traditional experimentalism for determining effectiveness we consider them misguided. We think their attitude stems from self-interest. We

don't see our own conflict of interest: Who is going to benefit immediately from the requirement of performing random trials? Us. We see stakeholders as having conflicts of interest, but not ourselves.

We can't ignore that we are part of a larger struggle for societal change. We need to acknowledge the ramifications of our actions and make sure the information we provide is used responsibly.

*Some may write off this article as self-righteous rambling, but that is a symptom of the problem—we think it's always others causing the problems.*

### Moving Forward—Building Better Relations

I have great hopes for our profession. Some may write off this article as self-righteous rambling, but that is a symptom of the problem—we think it's always others causing the problems. The problem of how to relate to consumers does exist. While many call for reflection, a symposium, or a special publication on the topic, I would suggest that we look more structurally. The first step is to recognize that we are accountable to the public. Accountability and respect go together. On this front we need large-scale changes, like voluntary certification or licensure.

The next step is to recognize that we want to have a relationship with evaluation consumers. We should think about how can we get along and mutually support each other's needs—and apply what we learned in kindergarten: to be nice, to share, to not call each other names, and to play fairly.

David Chavis  
President

Association for the Study and Development of Community  
312 S. Frederick Avenue  
Gaithersburg, MD 20877  
Tel: 301-519-0722  
Email: [dchavis@capablecommunity.com](mailto:dchavis@capablecommunity.com)

<sup>3</sup> Bakan, D. (1965). The mystery-mastery complex in contemporary society. *American Psychologist*, 20, 186–191.

## A conversation with **Ricardo Millett**

Ricardo Millett is a veteran philanthropist and evaluator and is president of the Woods Fund of Chicago. The foundation is devoted to increasing opportunities for less advantaged people and communities in the Chicago metropolitan area—including their opportunities to contribute to decisions that affect their lives. Prior to joining the Woods Fund, Dr. Millett was director of evaluation at the W. K. Kellogg Foundation. He has also held management positions at the United Way of Massachusetts Bay and the Massachusetts Department of Social Services. Dr. Millett has been a consistent and prominent voice on issues of diversity in evaluation and has been instrumental in developing solutions that will better enable evaluators to address and build capacity around diversity and multiculturalism.

**Q** How should we be thinking about diversity as it applies to evaluation?

**A** Evaluators are in the business of interpreting reality through the kind of information we capture. If we are good at what we do, our work is seen as legitimate and we influence programs and policies. My concern is that much evaluation is not accurately capturing the experiences of those who are affected by the programs and policies we inform. Conventional program evaluation often misses the kinds of data and experiences that can help to frame effective programs and policies, and this problem relates directly to how we approach diversity and multiculturalism in our profession.

Jennifer Greene, Rodney Hopson, and I recently wrote a paper about the need for evaluation to generate *authentic knowledge* about social programs and issues.<sup>1</sup> This is knowledge that captures and authentically represents the experiences and perspectives of people affected by these programs or issues—often individuals or communities of color. Generating authentic knowledge is about finding a way to make sure that evaluation is participatory and grounded, and collects and interprets data within real settings. It is not about capturing whether participants work well for a program, but whether a program works well for participants.

Consider the issue of public housing. Many cities have developed efforts to transfer low-income residents out of public housing or high-rise projects into affordable mixed-residential developments. Sounds like a good idea, right? But once we had these programs in place, we suddenly realized that there were problems with this approach. We were moving people out of low-income housing faster than we could find alternative housing. Some individuals had deeply entrenched problems that

made them hard to place. Some programs shut males out of the transition program because they didn't take into account non-traditional conceptions of family dynamics and structure. And the support services that were previously available suddenly were not available in new neighborhoods.

So families had better housing, but now they had all sorts of new problems. That suggests to me that the planning and evaluation that helped to design these programs did not capture and relate the authentic experiences of those who actually experienced them, and did not use those experiences to craft effective transition programs. That kind of shortsightedness is the difference between a conventional approach to evaluation and what I call a multicultural approach that respects and captures authentic knowledge and experience as part of the evaluation process.

If we are going to get better at capturing authentic experience, we need to look more carefully at who is doing the evaluation and at the approach being used. We must ask who—in terms of ethnicity, background, training, and experience—is doing the evaluation and how they are doing it.

I am not suggesting that capturing authentic experience necessarily requires an evaluator to be of the same class and ethnicity as the individuals in the program being evaluated, though those considerations are critical. But I am suggesting that evaluators have to possess the sensitivities, abilities, and capacity to see experiences within their context. If we don't, then we are likely to do damage by helping to sustain ineffective policies or strategies. If we understand them enough and are willing enough to dig into these experiences with our evaluation approach, then we are more likely to capture authentic experience.

If not, we risk helping to legitimize the negative characterization of people in poverty and the programs or policies that keep them poor. Capturing authentic experience requires a little humility and an understanding that a lot of our work is more art and sociology than hard science.

*Capturing authentic  
experience requires a little  
humility and an  
understanding that a lot of  
our work is more art and  
sociology than hard science.*

<sup>1</sup> Greene, J. C., Millett, R., & Hopson, R. (in press). Evaluation as a democratizing practice. In M. Braverman, N. Constantine, & J. K. Slater (Eds.), *Putting evaluation to work for foundations*. San Francisco, CA: Jossey-Bass.

**Q** How would you characterize the evaluation field's current stance on issues of diversity?

**A** Several years ago, when I was director of evaluation at the W. K. Kellogg Foundation, a number of colleagues of color and I started a diversity project that originated from the questions many foundation program officers had about why so few evaluators of color were receiving evaluation contracts.

We developed a directory to identify evaluators of color across the nation. But then we realized that if we wanted to address this issue seriously, we needed to do more than a directory. There simply were not enough evaluators of color available, or not enough evaluators with the capacity to work across cultural settings.

As a result, the American Evaluation Association (AEA) and the Dusquene University School of Education have been engaged in a joint effort to increase racial and ethnic diversity and capacity in the evaluation profession. The project is developing a "pipeline" that will offer evaluation training and internship opportunities (in foundations, nonprofits, and the government) for students of color from various social science disciplines.

Initially, success for this pipeline project will mean that we no longer have to scour the world to find evaluators of color. If the project is further funded and expanded, long-term success will mean that the courses and tools that have been developed will be institutionalized into the broader realm of evaluation training and professional development and made available to all evaluators, not just those of color. Eventually, approaches that help us capture authentic experience will become a legitimate part of the way the evaluation profession does business.

In the beginning this idea met with some resistance and defensiveness in the broader evaluation community. Questions about eligibility for internship participation and even the need for such an approach surfaced, along with the feeling that the notion of multicultural evaluation was not something that should be legitimated. This resistance has diminished over time, but it is something that the field as a whole must continue to struggle with. Now we are having more open dialogue about these issues, spurred in large part by the very active and vocal efforts of the Multiethnic Issues in Evaluation topical interest group within AEA.

**Q** What has improved over the past decade?

**A** Ten years ago, we—meaning evaluators of color—were isolated and frustrated that these issues about diversity in evaluation were not on anyone's radar. Ten years ago there weren't enough evaluators of color in leadership positions; and there weren't enough academicians and practitioners

for whom this issue resonated.

Ten years later, we have not just evaluators of color pushing this issue, we have a range of evaluators in leadership positions supporting it. The articulation of these concerns has become sharp, coherent, and effective in getting the attention of major stakeholders in the funding world and at academic institutions. The response has been much greater, and more foundations are willing and ready to take these issues on and build the capacity that the evaluation profession needs.

**Q** What should we be doing to make real and sustainable change on issues of diversity in evaluation?

**A** In addition to raising the profile of these issues, offering more education on approaches for capturing authentic experience, and increasing the number of evaluators of color, we should be paying attention to what evaluators in other countries are doing. The kind of evaluation that is participatory and captures authentic experience is almost standard in the third world. We have been slow in this country to learn and adapt.

Also, more often than not we accept and compromise the principles of truth for a contract. We offer and accept evaluation dollars that are less than what we need to get good and authentic information. We accept forced definitions of problems, and we don't push what we know to be true.

As evaluators we need to play out our responsibility to generate data that is as true as possible to interpreting the reality of people that are being served, and not legitimate the programs and policies that keep them from having a voice in and improving their own conditions.



Ricardo Millett

Julia Coffman, Consultant, HFRP  
Email: [julia\\_coffman@msn.com](mailto:julia_coffman@msn.com)

### Related Resources

Mertens, D. M. (1997). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.

Part of the American Evaluation Association, the Multiethnic Issues in Evaluation topical interest group's mission is to (1) raise the level of discourse on the role of people of color in the improvement of the theory, practice, and methods of evaluation, and (2) increase the participation of members of racial and ethnic minority groups in the evaluation profession. [www.obsidcomm.com/aea](http://www.obsidcomm.com/aea)

## A Decade of International Trends in Evaluation

*Craig Russon describes a decade of efforts to link a growing number of regional and national evaluation organizations into a worldwide community through the International Organization for Cooperation in Evaluation.*

Congratulations to *The Evaluation Exchange* on its 10-year anniversary. It is an interesting coincidence that the growth in the worldwide community of evaluators began at about the same time that *The Evaluation Exchange* began publication. Prior to 1995, there were only five regional and national evaluation organizations: American Evaluation Association (AEA), Australasian Evaluation Society (AES), Canadian Evaluation Society (CES), Central American Evaluation Association (ACE), and European Evaluation Society (EES). Today there are about 50! For a number of years, efforts have been made to create a loose coalition of these evaluation organizations.

These efforts date back to the 1995 international conference in Vancouver, British Columbia, sponsored by the American Evaluation Association and Canadian Evaluation Society. The theme of the conference was “Evaluation for a New Century—A Global Perspective.” Delegates from 50 countries attended the event and many came away thinking about evaluation in new ways. A couple of years later, a discussion took place on the EvalTalk listserv regarding the international nature of the profession. One of the principal issues discussed was the creation of a federation of national evaluation organizations.

As a result of that discussion, the International & Cross-Cultural Evaluation Topical Interest Group (I&CCE) convened a panel of six regional and national evaluation organization presidents. The Presidents Panel was a plenary session at the 1998 annual meeting of the AEA (Russon & Love, 1999). The purpose of the panel was to discuss the creation of a “worldwide community of evaluators.” One of the outcomes of the panel was the decision to move slowly ahead with this project. A proposal was developed and funding was obtained from the W. K. Kellogg Foundation (WKKF) to take the next step.

February 18–20, 1999, a residency meeting was held in Barbados, West Indies, to discuss the issues associated with creating this worldwide community (Mertens & Russon, 2000). The meeting was attended by the leaders of 15 regional and national evaluation organizations from around the world. Also in attendance were observers from WKKF, the University of the West Indies, the Caribbean Development Bank, and the UN Capital Development Fund. Through intense negotiations, the group was able to identify the purposes, organizational principles, and activities that would underpin the worldwide community. A drafting committee that represented the diverse nature of the group was selected to develop a charter for what would come to be called the International Organization for Cooperation in Evaluation (IOCE).

It took nearly a year for the charter to be endorsed by all of the organizations that were represented at the Barbados meeting. Then the charter was presented to the rest of the regional and national evaluation organizations around the world. With the sup-

*continued on page 19*

### International Evaluation Organizations

African Evaluation Association  
American Evaluation Association  
Association Comorienne de Suivi et Evaluation (Comoros)  
Associazione Italiana de Valuatazione (Italy)  
Australasian Evaluation Society (Australia and New Zealand)  
Bangladesh Evaluation Forum  
Botswana Evaluation Association  
Brazilian M&E Network  
Burundi Evaluation Network  
Canadian Evaluation Society  
Central American Evaluation Association  
Danish Evaluation Society  
Deutsche Gesellschaft für Evaluation (Denmark)  
Egyptian Evaluation Association  
Eritrea Evaluation Network  
Ethiopian Evaluation Association  
European Evaluation Society  
Finnish Evaluation Society  
Ghana Evaluators Association  
Ghana Evaluation Network  
International Program Evaluation Network (Russia/Newly Independent States)  
Israeli Association for Program Evaluation  
Japanese Evaluation Association  
Kenya Evaluation Association  
Korean Evaluation Association  
La Société Française de l'Evaluation (France)  
Malawi M&E Network  
Malaysian Evaluation Society  
Namibia Monitoring Evaluation and Research Network  
Nepal M&E Forum  
Network for Monitoring and Evaluation of Latin America and the Caribbean  
Nigerian Evaluation Association  
Programme for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean  
Reseau Malgache de Suivi et Evaluation (Malagasy)  
Reseau Nigerien de Suivi et Evaluation (Nigeria)  
Reseau Ruandais de Suivi et Evaluation (Rwanda)  
Société Québécoise d'Evaluation de Programme  
Société Wallonne de l'Evaluation et de la Prospective (Belgium)  
South African Evaluation Network  
Spanish Public Policy Evaluation Society (Spain)  
Sri Lanka Evaluation Association  
Swiss Evaluation Society  
Thailand Evaluation Network  
Ugandan Evaluation Association  
United Kingdom Evaluation Society  
Utvarderarna (Sweden)  
Zambia Evaluation Association  
Zimbabwe Evaluation Society



## An Update on University-Based Evaluation Training

*Molly Engle and James Altschuld describe results from their research on recent trends in university-based evaluation training.*

Preparing evaluators is an ongoing process and one that engages many individuals in universities, colleges, government agencies, and professional organizations. No two paths to the evaluation profession are the same. Reviewing the current opportunities for preparing evaluators allows us to see progress, identify where growth can and is occurring, and to preserve the profession's history.

Early in 2000, the American Evaluation Association (AEA) endorsed a project funded by the National Science Foundation (NSF) to update what we know about university-based evaluation training. In addition, the evaluation team was charged with determining what kind of professional development training was being offered. Assisted by Westat, a leading research corporation, we conducted two surveys, one for each type of training in question. Both surveys were international in scope. This article presents preliminary findings of the university-based survey.

First, a little history. In 1993, we conducted a similar survey<sup>1</sup> in which we defined a “program” as a curricular offering of two or more courses in sequence, specifically, “A program consists of multiple courses, seminars, practicum offerings, etc., designed to teach what the respondent considered to be evaluation principles and concepts.”<sup>2</sup> This statement made it possible to interpret “program” in a variety of ways, but it clearly excluded single-course programs.

At that time we identified a total of 49 programs. Thirty-eight were based in the United States, a decrease from the previous study, conducted in 1986,<sup>3</sup> which found 44. We also identified 11 programs internationally, all in Canada or Australia. Three programs were in government agencies and there was one nontraditional program, which did not exist in 1986. It is important to note that of these 49 programs, only one half (25) had the word “evaluation” in their official title, limiting the visibility of the others.

The process we used for the current survey was similar to that used in 1993. We distributed a call for nominations through various listservs including Evaltalk, Govteval, and XCeval, as well as through personal communication and general solicitation at AEA's annual meeting. We developed a sam-

pling frame of 85 university-based programs and 57 professional development offerings (not discussed here). A unique aspect of the current survey that differed from the previous surveys was that NSF requested we examine whether any training programs focused on science, technology, math, or engineering. In addition, the AEA Building Diversity Initiative, an ad hoc committee, requested that we develop a mechanism to determine the extent of training programs in minority-serving institutions such as the Historically Black Colleges and Universities (often the 1890 land-grant institutions), Hispanic Serving Institutions, and Tribal Institutions (the 1994 land-grant institutions). The sampling frame for minority-serving institutions was developed separately and returned a list of 10 schools offering individual courses.

The preliminary results from the current study show that the face of university-based evaluation training has once again changed. The total number of programs has decreased from 49 in 1993 to 36—26 United States programs and 10 international programs. One reason for this decrease could be that senior evaluation leaders are retiring from their academic lives. Often these programs were the passion of a single individual who developed a collaborative and interdisciplinary program. We have not yet begun to see the next generation of university-based programs led by passionate young faculty.

Of those 36 institutions responding, 22 (61%) have “evaluation” in their formal title. The lack of a recognizable program title remains problematic for the future of the profession. If individuals are unable to quickly locate training opportunities in evaluation, they will be more likely to choose a different course of study. This could lead to a further reduction of university-based programs due to low enrollments, to an increase in alternative training opportunities, or to some hybrid approach to entry into the profession.

*The preliminary results from the current study show that the face of university-based evaluation training has once again changed.*

*The total number of programs has decreased from 49 in 1993 to 36.*

**Molly Engle, Ph.D.**  
Associate Professor of Public Health  
College of Health and Human Sciences  
Oregon State University  
307 Ballard Extension Hall  
Corvallis, OR 97331  
Tel: 541-737-4126  
Email: [molly.engle@oregonstate.edu](mailto:molly.engle@oregonstate.edu)

**James W. Altschuld, Ph.D.**  
Professor of Education  
The Ohio State University  
310B Ramseyer  
29 W. Woodruff Avenue  
Columbus, OH 43210  
Tel: 614-292-7741  
Email: [altschuld.1@osu.edu](mailto:altschuld.1@osu.edu)

<sup>1</sup> Altschuld, J. W., & Engle, M. (Eds.). (1994). The preparation of professional evaluators: Issues, perspectives, and programs. *New Directions for Program Evaluation*, 62.

<sup>2</sup> Altschuld, J. W., Engle, M., Cullen, C., Kim, I., & Macce, B. R. (1994). The directory of evaluation training programs. *New Directions for Program Evaluation*, 62, 72.

<sup>3</sup> May, R. M., Fleischer, M., Schreier, C. J., & Cox, G. B. (1986). Directory of evaluation training programs. In B. G. Davis (Ed.), *New Directions for Program Evaluation*, 29, 71–98.

### Using Narrative Methods to Link Program Evaluation and Organization Development

*Charles McClintock, Dean of the Fielding Graduate Institute's School of Human and Organization Development, illustrates how narrative methods can link a program's evaluation and its organization development.*

The field of program evaluation has evolved over the past half century, moving from focusing primarily on research methods to embracing concepts such as utilization, values, context, change, learning, strategy, politics, and organizational dynamics. Along with this shift has come a broader epistemological perspective and wider array of empirical methods—qualitative and mixed methods, responsive case studies, participatory and empowerment action research, and interpretive and constructivist versions of knowledge.

Still, evaluation has remained an essentially empirical endeavor that emphasizes data collection and reporting and the underlying skills of research design, measurement, and analysis. Related fields, such as organization development (OD), differ from evaluation in their emphasis on skills like establishing trusting and respectful relationships, communicating effectively, diagnosis, negotiation, motivation, and change dynamics. The future of program evaluation should include graduate education and professional training programs that deliberately blend these two skill sets to produce a new kind of professional—a scholar-practitioner who integrates objective reflection based on systematic inquiry with interventions designed to improve policies and programs (McClintock, 2004).

Narrative methods represent a form of inquiry that has promise for integrating evaluation and organization development. Narrative methods rely on various forms of storytelling that, with regard to linking inquiry and change goals, have many important attributes:

1. Storytelling lends itself to participatory change processes because it relies on people to make sense of their own experiences and environments.
2. Stories can be used to focus on particular interventions while also reflecting on the array of contextual factors that influence outcomes.
3. Stories can be systematically gathered and claims verified from independent sources or methods.
4. Narrative data can be analyzed using existing conceptual frameworks or assessed for emergent themes.
5. Narrative methods can be integrated into ongoing organizational processes to aid in program planning, decision making, and strategic management.

The following sketches describe narrative methods that have somewhat different purposes and procedures. They share a focus on formative evaluation, or improving the program during its evaluation, though in several instances they can contribute to summative assessment of outcomes. For purposes of comparison, the methods are organized into three groups: those that

are relatively structured around success, those whose themes are emergent, and those that are linked to a theory of change.

#### Narratives Structured Around Success

Dart and Davies (2003) propose a method they call the *most significant change* (MSC) technique and describe how it was applied to the evaluation of a large-scale agricultural extension program in Australia. This method is highly structured and designed to engage all levels of the system from program clients and front-line staff to statewide decision makers and funders, as well as university and industry partners. The MSC process involves the following steps:

1. Identify domains of inquiry for storytelling (e.g., changes in decision-making skills or farm profitability).
2. Develop a format for data collection (e.g., story title, what happened, when, and why the change was considered significant).
3. Select stories by voting at multiple levels (e.g., front-line staff, statewide decision makers and funders) on those accounts that best represent a program's values and desired outcomes.
4. Conduct a content analysis of all stories (including those not selected in the voting) in relation to a program logic model.<sup>1</sup>

As described by Dart and Davies (2003), one of the most important results of MSC was that the story selection process surfaced differing values and desired outcomes for the program. In other words, the evaluation storytelling process was at least as important as the evaluation data in the stories. In addition, a follow-up case study of MSC revealed that it had increased involvement and interest in evaluation, caused participants at all levels to understand better the program outcomes and the dynamics that influence them, and facilitated strategic planning and resource allocation toward the most highly valued directions. This is a good illustration of narrative method linking inquiry and OD needs.

A related narrative method, structured to gather stories about both positive and negative outcomes, is called the *success case method* (Brinkerhoff, 2003 [A review of this book is available in this issue on page 16. —Ed.]). The method has been most frequently used to evaluate staff training and related human resource programs, although conceptually it could be applied to other programs as well.

This method has two phases. A very short email or mail survey is sent to all program participants to identify those for whom the training made a difference and those for whom it did not. Second, extreme cases are selected from those two ends of the success continuum and respondents are asked to tell stories about both the features of the training that were or were not helpful as well as other organizational factors that facilitated or

<sup>1</sup> A logic model illustrates how the program's activities connect to the outcomes it is trying to achieve.

impeded success (e.g., support from supervisors and performance incentives). Based on the logic of journalism and legal inquiry, independent evidence is sought during these storytelling interviews that would corroborate the success claims.

The purpose of the success case method is not just to evaluate the training, but to identify those aspects of training that were critical—alone or in interaction with other organizational factors. In this way, the stories serve both to document outcomes, but also to guide management about needed organizational changes that will accomplish broader organizational performance goals. Kibel (1999) describes a related success story method that involves more complex data gathering and scoring procedures and that is designed for a broader range of human service programs.

### Narratives With Emerging Themes

A different approach to narrative methods is found within *qualitative case studies* (Costantino & Greene, 2003). Here, stories are used to understand context, culture, and participants' experiences in relation to program activities and outcomes. As with most case studies, this method can require site visits, review of documents, participant observation, and personal and telephone interviews. The authors changed their original practice of summarizing stories to include verbatim transcripts, some of which contained interwoven mini stories. In this way they were able to portray a much richer picture of the program (itself an intergenerational storytelling program) and of relationships among participants and staff, and they were able to use stories as a significant part of the reported data.

Nelson (1998) describes a similar approach that uses both individual and *group storytelling* in evaluating youth development and risk prevention programs. The individual stories elicit participant experiences through a series of prompts, while the group stories are created by having each group member add to a narrative about a fictitious individual who participates in the program and then has a set of future life outcomes. Group storytelling is a means of getting at experiences an individual is reluctant to claim or at material that might not be accessible to conscious thought.

Both of these approaches can result in wide differences in the quality and detail of the stories. Especially with group storytelling, the narrative can become highly exaggerated. The point of narrative in these instances is not so much to portray factual material as it is to convey the psychological experience of being in the program. Analysis can take many forms, depending on the conceptual framework or evaluation contract, and can include thematic coding, verbatim quotes, and narrative stories as the substance of the analysis.

### Narratives Linked to a Theory of Change

The previous uses of narrative emphasize inquiry more than OD perspectives. *Appreciative inquiry* (AI) represents the opposite emphasis, although it relies heavily on data collection and analysis (Barrett & Fry, 2002). The AI method evolved over time within the OD field as a form of inquiry designed to identify potential for innovation and motivation in organizational groups.

AI is an attempt to move away from deficit and problem-solving orientations common to most evaluation and OD work and move toward “peak positive experiences” that occur within organizations. AI uses explicitly collaborative interviewing and narrative methods in its effort to draw on the power of social constructionism to shape the future. AI is based on social constructionism's concept that what you look for is what you will find, and where you think you are going is where you will end up.

The AI approach involves several structured phases of systematic inquiry into peak experiences and their causes, along with creative ideas about how to sustain current valued innovations in the organizational process. Stories are shared among stakeholders as part of the analysis and the process to plan change. AI can include attention to problems and can blend with evaluation that emphasizes accountability, but it is decidedly effective as a means of socially creating provocative innovations that will sustain progress.

This brief overview of narrative methods shows promise for drawing more explicit connections between the fields of program evaluation and OD. In addition, training in the use of narrative methods is one means of integrating the skill sets and goals of each profession to sustain and improve programs.

**Charles McClintock, Ph.D.**

Dean

School of Human and Organizational Development

Fielding Graduate Institute

2112 Santa Barbara Street

Santa Barbara, CA 93105

Tel: 805-687-1099

Email: [cmclintock@fielding.edu](mailto:cmclintock@fielding.edu)

### References

- Barrett, F., & Fry, R. (2002). Appreciative inquiry in action: The unfolding of a provocative invitation. In R. Fry, F. Barrett, J. Seiling, & D. Whitney (Eds.), *Appreciative inquiry and organizational transformation: Reports from the field*. Westport, CT: Quorum Books.
- Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not*. San Francisco: Berrett-Koehler Publishers.
- Costantino, R. D., & Greene, J. C. (2003). Reflections on the use of narrative in evaluation. *American Journal of Evaluation*, 24(1), 35–49.
- Dart, J., & Davies, R. (2003). A dialogical, story-based evaluation tool: The most significant change technique. *American Journal of Evaluation*, 24(2), 137–155.
- Kibel, B. M. (1999). *Success stories as hard data: An introduction to results mapping*. New York: Kluwer/Plenum.
- McClintock, C. (2004). The scholar-practitioner model. In A. DiStefano, K. E. Rudestam, & R. J. Silverman (Eds.), *Encyclopedia of distributed learning* (pp. 393–396). Thousand Oaks, CA: Sage.
- Nelson, A. (1998). *Storytelling for prevention*. Evergreen, CO: The WHEEL Council.

## The Success Case Method: Finding Out What Works

*From time to time, we will print reviews of noteworthy new books in the evaluation field. In this, our first review, Tezeta Tulloch from Harvard Family Research Project takes a look at Robert Brinkerhoff's The Success Case Method, a handy, accessible guide for both experienced evaluators and novices.*

**T**he *Success Case Method*, a new book by Robert Brinkerhoff, offers an engaging discussion of the success case method (SCM) in evaluation with step-by-step instructions on how to implement it. Brinkerhoff describes the method as a blend of storytelling and rigorous evaluation methods. SCM is, in short, a relatively quick and cost-effective way of determining which components of an initiative are working and which ones are not, and reporting results in a way that organization leaders can easily understand and believe.

The method uses the stories of individuals participating in an initiative to investigate and understand the roots of their successes and proceeds from the premise that small successes can lead to greater ones. If only five out of 50 participants achieve marked success, it follows that a detailed study of what these five are doing could yield instructive results for those who are struggling.

Picking out the best (and sometimes, worst) accounts, verifying and then documenting them is the core of the success case method. This confirmation step is what distinguishes this approach from a simple cataloguing of positive accounts or anecdotes. Cases selected to represent an initiative's strengths undergo rigorous scrutiny, passing muster only when respondents' claims can be concretely confirmed. "A success story," says Brinkerhoff, "is not considered valid and reportable until [it can] stand up in court."

One of the SCM's strengths is assessing what Brinkerhoff calls "soft" interventions, such as communication and other interpersonal capabilities that are generally difficult to measure. In one example, he cites an emotional intelligence training program for sales representatives obligated to cold-call customers. A success case assessment of the trainees found strong evidence that their use of trained skills decreased their fears of rejection, which led to an increase in successful calls completed, which in turn led to increased sales.

The method can also be used to estimate quickly return on investment by comparing an estimate of the dollar value of successful results with the cost of implementing the program for the participants that achieved those results. Additionally, SCM can help calculate a program's "unrealized value"—the additional benefits a program can achieve if a greater number of people were to use the same methods employed by its more successful users. This estimate is especially useful in helping program proponents make a "business case" for improving a program that is currently underachieving.

SCM can be especially useful when integrated into a bigger effort to engineer organizational change. By quickly identifying the aspects of an initiative that are bringing about positive results, the strategy can help organizations hone in on which ele-

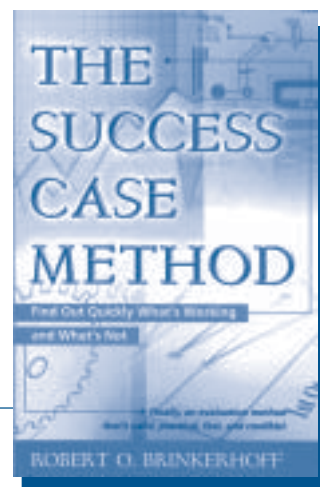
ments to nurture and which ones to discard before too many resources have been invested in a failing program. Alternatively, the method can be used to salvage useful parts of a program already slated for termination.

Brinkerhoff is careful to acknowledge the limits of this approach. In his discussion on conducting surveys, he notes that some ("single purpose") surveys can take the form of a single, simple question: "Who among your staff is the most up-to-date on current office affairs?" Theoretically, anyone can handle this sort of data gathering. For those who need to elicit a wider,

### Featured Book

*The Success Case Method: Find Out Quickly What's Working and What's Not.*

Robert O. Brinkerhoff.  
192 pages. (2003).  
San Francisco: Berrett-Koehler Publishers, Inc.  
\$24.95



more diverse ("multipurpose") range of information, Brinkerhoff recommends seeking expert advice on what can be an extremely complicated procedure.

No matter what the level of sophistication, "one major and enduring challenge for any sort of organizational analysis is trying to get key people and groups to pay attention to findings." While there is no way to ensure a unanimously positive response, evaluators have at their disposal various reporting tactics for piquing interest. These include live presentations, video dramatizations, various report formats, and workshops for key stakeholders to discuss and apply findings. Inviting some stakeholders to participate in data collection is another way to promote investment in an evaluation's outcomes, Brinkerhoff suggests. The book, though, offers little in the way of how those without formal experience can prepare themselves to participate actively in "data collection and analysis activities."

Brinkerhoff's book is bolstered by the rich selection of initiatives he draws on as examples and the charts and diagrams that illustrate essential steps—the order in which to proceed, the kinds of questions to ask, and so forth. What *The Success Case Method* profits from most is a lively, informal writing style that should appeal to a broad spectrum of managers, organizers, and other potential change agents, as well as evaluation beginners and experts.

Tezeta Tulloch, Publications Assistant, HFRP  
Email: [tezeta\\_tulloch@harvard.edu](mailto:tezeta_tulloch@harvard.edu)



## Evaluation Strategies to Support Policymaking in Learning Technology

*Geneva Haertel and Barbara Means of SRI International describe how evaluators and policymakers can work together to produce “usable knowledge” of technology’s effects on learning.*

Evaluating technology’s effect on learning is more complicated than it appears at first blush. Even defining what is to be studied is often problematic. Educational technology funds support an ever-increasing array of hardware, software, and network configurations that often are just one aspect of a complex intervention with many components unrelated to technology. Since it is the teaching and learning mediated by technology that produces desired results rather than the technology itself, evaluation should examine the potential influence of teachers, students, and schools on learning.

Understandably, policymakers tend to leave the evaluation of educational technology to evaluation professionals. But we believe policymakers and evaluators should work in tandem—by collaborating they can avoid the intellectual stumbling blocks common in this field. While evaluators bring specialized knowledge and experience, they are not in the best position to set priorities among competing questions. This is the realm of policymakers, who need to think carefully about the kinds of evidence that support their decisions.

In a recent volume (Means & Haertel, 2004), we identify six steps evaluators and policymakers can take to produce more useful evaluations of learning technologies.

1. *Clarify evaluation questions.* The language of the No Child Left Behind Act often is construed to mean that the only relevant question is technology’s impact on achievement—an important question, but not the only one local policymakers care about. In some cases implementation of a technology (say Internet access for high schools) is a foregone conclusion, and instead policymakers may need to address an issue such as how best to integrate the technology with existing courses.

2. *Describe technology-supported intervention.* Evaluators, policymakers, and other stakeholders should work together to develop a thorough description of the particular technology-supported intervention in question. A theory of change (TOC) approach would specify both the outcomes the intervention is expected to produce and the necessary conditions for attaining them.

3. *Specify context and degree of implementation.* Evaluators and policymakers should identify both those served by the intervention and those participating in the evaluation. At this point, they should also specify the degree to which the intervention has been implemented. They can pose questions such as (1) What degree of implementation has occurred at the various sites? and (2) Have teachers had access to the training they need to use the technology successfully?

Answers will enable evaluators to advise policymakers on whether to conduct a summative evaluation or an implementation evaluation. Some informal field observations of the tech-

nology can also be helpful at this point. This is the stage where the original purpose of the evaluation is confirmed or disconfirmed.

4. *Review student outcomes.* The outcomes measured will be those targeted in the TOC. Evaluators can generate options for the specific methods and instruments for measuring outcomes. Some technologies aim to promote mastery of the kinds of discrete skills tapped by most state achievement tests; others support problem-solving skills rarely addressed by achievement tests. A mismatch between the learning supported by an intervention and that measured as an outcome can lead to erroneous conclusions of “no effect.”

Evaluators and policymakers will need to prioritize outcomes, picking those that are most valued and for which information can be collected at a reasonable cost.

5. *Select evaluation design.* The choice of evaluation design requires both the expertise of evaluators and policymaker buy-in. True (random-assignment) experiments, quasi-experiments, and case studies are all appropriate designs for some research questions. While federal legislation promotes the use of true experiments, it is easier to conduct experiments on shorter term, well-defined interventions than on longer term or more open-ended interventions.

6. *Stipulate reporting formats and schedule.* Policymakers and evaluators should agree in advance of data collection on the nature, frequency, and schedule of evaluation reports. Reporting formats should make sense to a policy audience and provide data in time to inform key decisions.

To produce “usable knowledge,” or professional knowledge that can be applied in practice (Lagemann, 2002), we call for (1) evaluations that address the questions that policymakers and practitioners care about, (2) integration of local understandings produced by evaluator-policymaker partnerships with disciplinary knowledge, and (3) use of evaluation findings to transform practice.

### References and Related Resources

Haertel, G. D., & Means, B. (2003). *Evaluating educational technology: Effective research designs for improving learning*. New York: Teachers College Press.

Lagemann, E. C. (2002). *Usable knowledge in education*. Chicago: Spencer Foundation. [www.spencer.org/publications/index.htm](http://www.spencer.org/publications/index.htm)

Means, B., & Haertel, G. D. (2004). *Using technology evaluation to enhance student learning*. New York: Teachers College Press.

**Geneva D. Haertel, Ph.D.**  
Senior Educational Researcher  
Tel: 650-859-5504  
Email: [geneva.haertel@sri.com](mailto:geneva.haertel@sri.com)

**Barbara Means, Ph.D.**  
Center Director  
Tel: 650-859-4004  
Email: [barbara.means@sri.com](mailto:barbara.means@sri.com)

Center for Technology in Learning  
SRI International  
333 Ravenswood Ave., BN354  
Menlo Park, CA 94025

## Transforming School Culture Through the Arts

*Programs that use the performing and visual arts as their core intervention play an increasing role in addressing various social issues and problems, yet little information has been shared about how to evaluate them. This article leads off The Evaluation Exchange's commitment to assist in such information sharing.*

Imagine artists of diverse races and ages leading a classroom of children in a tribal yell, or guiding the children in a human chain, as they weave through the room making music with ancient and handmade instruments. This is the everyday work of artists in the Tribal Rhythms® company, a program of the Cooperative Artists Institute (CAI).

CAI is a multicultural nonprofit in Boston that uses the performing and visual arts to help schools and communities solve problems, especially those relating to community and family fragmentation. CAI created the Partnership for Whole School Change (the Partnership), a collaboration of CAI; Troubador, Inc.; Lesley University's Center for Peaceable Schools; and three elementary schools in Boston—Charles Taylor, Louis Agassiz, and Warren Prescott.

The Partnership is based on the belief that to improve school performance, communities need to create a school culture that has a positive effect on their children's behavior. To help achieve this transformation, the Partnership uses a range of strategies grounded in cultural anthropology. Tribal Rhythms is one of the Partnership's core strategies.

Partnership artist-educators use Tribal Rhythms to support schools in developing and implementing their school climate strategy. The program uses the themes of tribe, group building, and the arts to create nurturing, socially inclusive learning environments or "learning tribes" in classrooms and schools. Partnership artist-educators introduce the program with the Tribal Rhythms celebration, a highly participatory experience in which children play drums and other handcrafted instruments and act in dramatic stories. The celebration peaks when children help the artist-educator describe a strange and scary sighting by performing the "Dance of the Mysterious Creature." Afterward, teachers and artist-educators implement a series of lessons that incorporate dance, drama, and visual arts activities that reinforce the learning tribe concept and foster self-control, inclusiveness, and the values of caring, cooperation, and respect.

The goal is for students to see themselves as creators of culture as they develop a shared sense of community through their tribal ceremonies (e.g., tribal yells, signs, and council circles). By placing human relationships at the center of the instructional strategy, learning tribes promote an environment where teachers can spend more time teaching and less time preparing students to learn.

The Partnership integrates many of its programs' concepts into its evaluation. The evaluation recognizes the importance of using a participatory and team-based approach, employing teachers, Partnership service providers, and evaluators in the design of evaluation instruments, data collection, and interpretation. To improve the evaluation's validity, the project has been assessed from multiple perspectives (e.g., children, teachers, and artist-educators), using multiple methods: interviews with children in their Tribal Rhythms council circles, interviews with school staff and artist-educators, school-staff questionnaires, and student surveys.

One key evaluation component is a school climate survey administered to children to assess their feelings about their school and classroom over time. The idea is to determine whether the students' perceptions of school climate are changing, how social and antisocial behaviors are changing, and the teachers and artist-educators' roles—or lack thereof—in promoting these changes.

Partnership artist-educators teamed with first through fifth grade teachers to administer pre and post surveys in 16 classrooms in all three schools. To increase the survey's validity, artist-educators first led the younger children in a movement activity designed to help them better understand gradations, a concept needed to answer the survey questions. In one activity, children were asked how much energy they had that day. Hands low to the floor meant low energy, hands at the waist meant moderate energy, and hands over their heads meant high energy.

Although the evaluation is still in progress, results to date are intriguing in that they show that, over the course of the year, students generally felt positive about their classroom and school in the pre survey and negative in the post survey. However, their feelings about their learning tribe remained positive overall.

As the evaluation's purpose is both formative and summative, the evaluators are discussing the results with teachers and other school staff to gather their interpretations and to inspire future strategy. The goal is not just to create a learning culture in the three schools' classrooms, but also to have the evaluation contribute to a learning culture in the Partnership as a whole, where results can be discussed in an inclusive way, without apprehension, and decisions about their implications made through consensus.

**J. Curtis Jones**  
Partnership Coordinator  
Partnership for Whole School Change  
311 Forest Hills Street  
Jamaica Plain, MA 02130  
Tel: 617-524-6378  
Email: cai@tribal-rhythms.org

*To increase the survey's validity, artist-educators first led the younger children in a movement activity designed to help them better understand gradations, a concept needed to answer the survey questions.*

## International Trends

*continued from page 12*

port of the worldwide community of evaluators, a second proposal was developed and additional funding was obtained from WKKF. Members of the drafting committee met March 8–10, 2002, in the Dominican Republic and formed an Organizing Group to plan the inaugural assembly of the IOCE. Among the principal issues that the Organizing Group discussed during the meeting were participation, format, agenda, advanced processes, location, and secretariat.

These efforts culminated in the inaugural assembly of the IOCE. The event took place March 28–30, 2003, in Lima, Peru. It was attended by 40 leaders from 26 regional and national evaluation organizations from around the world (Russon & Love, 2003). An important objective that was achieved during the inaugural assembly was that a provisional constitution was endorsed. The constitution sets out the mission and organizational principles of the IOCE. The mission of the IOCE is:

... to legitimate and strengthen evaluation societies, associations, or networks so that they can better contribute to good governance and strengthen civil society. It will build evaluation capacity, develop evaluation principles and procedures, encourage the development of new evaluation societies and associations or networks, undertake educational activities that will increase public awareness of evaluation, and seek to secure resources for cooperative activity. It will be a forum for the exchange of useful and high quality methods, theories, and effective practice in evaluation.

Despite its short life, the influence of the IOCE is already being felt. Several new Latin American evaluation organizations were formed in advance of the IOCE inaugural assembly (e.g., Brazil, Colombia, and Peru). These organizations have joined with the Programme for Strengthening the Regional Capacity for Evaluation of Rural Poverty Alleviation Projects in Latin America and the Caribbean (PREVAL) and the Central American Evaluation Association (ACE). Together, they launched a regional organization called the Network for Monitoring and Evaluation of Latin America and the Caribbean (ReLAC) in Sao Paulo, Brazil, during September 2003. ReLAC and its member organizations are all affiliated the IOCE. Through the IOCE, a system of evaluation organizations is being created that will help us reinterpret the work that we have done in the past decade. It may suggest some ways that we should do our current work. And lastly, it may provide some insights into where we want to take this work in the future.

## References

- Mertens, D., & Russon, C. (2000). A proposal for the International Organization for Cooperation in Evaluation. *The American Journal of Evaluation*, 21(2), 275–284.
- Russon, C., & Love, A. (2003). *The inaugural assembly of the International Organization for Cooperation in Evaluation: The realization of a utopian dream*. Manuscript submitted for publication.

Russon, C., & Love, A. (Eds.). (1999). *Creating a worldwide evaluation community* (Occasional Paper Series, No. 15). Kalamazoo: The Evaluation Center, Western Michigan University. [Available at [www.wmich.edu/evalctr/pubs/ops/ops15.html](http://www.wmich.edu/evalctr/pubs/ops/ops15.html).]

**Craig Russon**  
Evaluation Manager  
W. K. Kellogg Foundation  
One Michigan Avenue East  
Battle Creek, MI 49017  
Tel: 269-968-1611  
Email: [craig.russon@wkcf.org](mailto:craig.russon@wkcf.org)

## New Resources From HFRP

We recently added a **bibliography of the evaluations of out-of-school time programs we are currently tracking nationwide** to the out-of-school time (OST) section of our website. Our bibliography contains about 230 programs and, if they are available online, links to their evaluation reports. [www.gse.harvard.edu/hfrp/projects/afterschool/bibliography/index.html](http://www.gse.harvard.edu/hfrp/projects/afterschool/bibliography/index.html)

We will continue to add entries to the bibliography as we learn about additional program evaluations. To be notified when we make updates, sign up for our out-of-school time updates email at [www.gse.harvard.edu/hfrp/subscribe.html](http://www.gse.harvard.edu/hfrp/subscribe.html).

The Family Involvement Network of Educators (FINE) has released a few new publications now available on the FINE website:

- The Fall 2003 edition of *FINE Forum*, FINE's biannual e-newsletter, asks how we can renew parent-teacher relations. [www.gse.harvard.edu/hfrp/projects/fine/fineforum/forum7/director.html](http://www.gse.harvard.edu/hfrp/projects/fine/fineforum/forum7/director.html)
- *Transforming Schools Through Community Organizing: A Research Review* examines current research on community organizing for school reform. It looks at how community organizing differs from traditional parent involvement activities, outlines the characteristic strategies used to engage parents in organizing efforts, and describes the outcomes of these efforts. [www.gse.harvard.edu/hfrp/projects/fine/resources/research/lopez.html](http://www.gse.harvard.edu/hfrp/projects/fine/resources/research/lopez.html)
- *Bridging Multiple Worlds: A Pathway Model for Minority Youth* provides a model for increasing opportunities for minority youth to succeed in school, attain college degrees, and prepare for careers. It describes how university research projects and cultural community programs can partner to support youth development. [www.gse.harvard.edu/hfrp/projects/fine/resources/research/minority.html](http://www.gse.harvard.edu/hfrp/projects/fine/resources/research/minority.html)

FINE membership is free of charge and is a great way to stay informed about family involvement and connect with others in the field. To join go to [www.gse.harvard.edu/hfrp/projects/fine/joinfine.html](http://www.gse.harvard.edu/hfrp/projects/fine/joinfine.html).

## Where We've Been, Where We're Going

continued from page 6

### On Evaluation and Philanthropy: Evaluation in a New Gilded Age

John Bare

Director of Planning and Evaluation, John S. and James L. Knight Foundation, Miami, Florida

If we're in a new Gilded Age, it's distinguished not by any new human frailty. Mark Twain's 19th century observation that man's chief end is to get rich, "dishonestly if we can, honestly if we must," certainly pertains in the 21st. What's different are management and evaluation practices that help us construct falsely precise measures in order to allocate resources to our liking.

Today's "corporate carnage," as *The Wall Street Journal* puts it, lays bare the myth that Fortune 500 management, and evaluation, will deliver philanthropy from the wilderness. Philanthropy has been urged to adopt practices that have contributed to, or at least made possible, Fortune 500 thievery. Adopted by governments, these practices gave us the Houston dropout scandal. No longer protected by tenure, principals ordered to make dropouts disappear—or else—reacted as rationally as Wall Street executives: They reported phony numbers.

Yet promoters keep hawking management and evaluation games that promise relief from hard-nosed questions of validity, internal and external. And I'll be damned if we aren't biting. Like a sucker on the carnival midway, philanthropy's booby prize is a cluster of pint-sized tables and graphics, called a "dashboard" for its mimicry of the gauge displays in cars. This innovation satisfies foundation trustees who refuse more than a page of explanation about knotty social change strategies.

The most promising remedies sound like riddles. To reject single-minded claims of measurement certainty does not require us to also reject the obligation to demonstrate value to society.

Ducking traps at the other extreme, we can value results without devaluing process. Both matter—*what* we accomplish and *how* we accomplish it—because values matter. When one man gets rich by stealing and another by hard work, the only thing separating them is *how* they got it done. The *how* matters, but only to the degree that it's connected to the *what*.

Wise voices are rising up. Michigan psychology professor Karl Weick tells *Harvard Business Review* that effective organizations "refuse to simplify reality." These "high-reliability organizations," or HROs, remain "fixed on failure. HROs are also fiercely committed to resilience and sensitive to operations."<sup>28</sup> Daniel Kahneman, the Princeton psychology professor who won the 2002 Nobel Prize in economics, explains in *Harvard Business Review* how an "outside view" can counter internal biases. Without it, Kahneman's "planning fallacy" takes hold, giving us "decisions based on delusional optimism."<sup>29</sup>

Delusions swell expectations, which in turn ratchet up pressure to cook the numbers, as illustrated by *USA Today's* item on Coca-Cola whistle-blower Matthew Whitley: "Just before midnight at the end of each quarter in 2002, Whitley alleges, fully loaded Coke trucks 'would be ordered to drive about two feet away from the loading dock' so the company could book 'phantom' syrup sales as part of a scheme to inflate revenue by tens of millions of dollars."<sup>30</sup>

Embracing the same management and evaluation practices, philanthropy will be ripe for the same whistle-blowing. Salvation lies in the evaluation paradox. Distilled, it is this: Our only hope for doing well rests on rewarding news about and solutions for whatever it is we're doing poorly.

<sup>28</sup> Weick, K. E., & Coutu, D. L. (2003, April). Sense and reliability: A conversation with celebrated psychologist Karl E. Weick. *Harvard Business Review*, 84–90.

<sup>29</sup> Lovallo, D., & Kahneman, D. (2003, July). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review*, 56–63.

<sup>30</sup> Smith, E. B. (2003, May 21). Coke investigates internal fraud allegations. *USA Today*, p. 1B.



Harvard Family Research Project  
Harvard Graduate School of Education  
3 Garden Street  
Cambridge, Massachusetts 02138  
hfrp@gse.harvard.edu  
www.hfrp.org

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

