5

Intended Process Uses

Impacts of Evaluative Thinking and Experiences

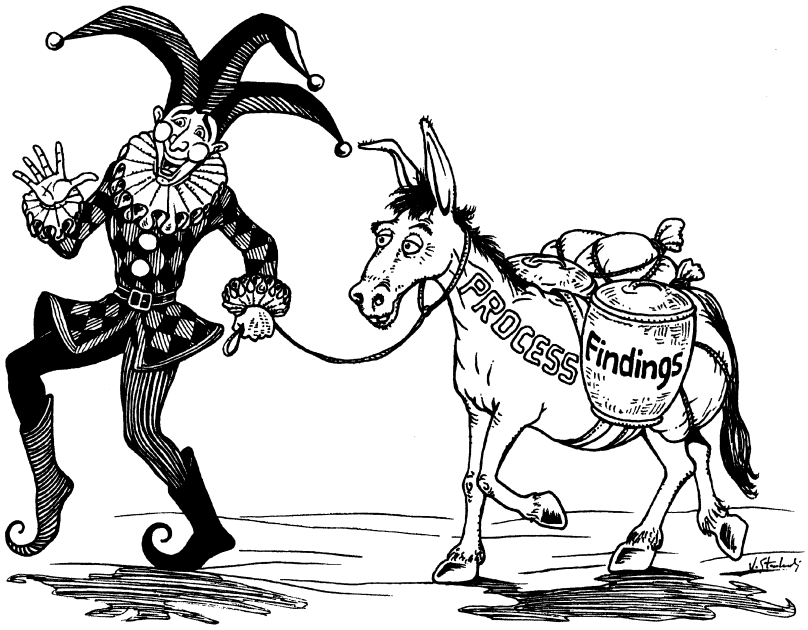
Utility is in the eye of the user.

—Halcolm

The Medium Is the Message

In the past, the search for use has often been conducted like the search for contraband in the famous Sufi story about Nasrudin the smuggler.

> *Nasrudin used to take his donkey across a frontier every day with the panniers loaded with straw. Since he admitted to being a smuggler, when he trudged home every night, the frontier guards searched him carefully. They searched his person, sifted the straw, steeped it in water, and even burned it from time to time. Meanwhile, he was becoming visibly more and more prosperous.*
>
> *Eventually, he retired to another country, very wealthy. Years later, one of the customs officials encountered him there. "You can tell me now, Nasrudin," he said. "Whatever was it that you were smuggling, that we could never catch you at?"*
>
> *"Donkeys," replied Nasrudin grinning (adapted from Shah 1964:59).*



Process as Outcome

In this chapter, we'll consider ways in which being engaged in the processes of evaluation can be useful quite apart from the findings that may emerge from those processes. Reasoning processes are evaluation's donkeys; they carry the load. Reasoning like an evaluator and operating according to evaluation's values have impacts.

When I refer to "process use," then, I mean using the logic, employing the reasoning, and being guided by the values that inform our practice. One way of thinking about process use is to recognize that evaluation constitutes a culture, of sorts. We, as evaluators, have our own values, our own ways of thinking, our own language, and our own reward system. When we engage other people in the evaluation process, we are providing them with a cross-cultural experience. They often experience evaluators as imperialistic, that is, as imposing the evaluation culture on top of their own values and culture—or they may find the cross-cultural experience stimulating and friendly. But in either case, and all the spaces in-between, it is a cross-cultural interaction. Those new to the evaluation culture may need help and facilitation in coming to view the experience as valuable. One of the ways I sometimes attempt to engage people in the value of evaluation is to suggest that they may reap personal and professional benefits from learning how to operate in an evaluation culture. Many funders are immersed in that culture. Knowing how to speak the language of evaluation and conceptualize programs logically are not inherent goods, but can be instrumentally good in helping people get the things they want, not least of all, to attract resources for their programs and make their work more effective. They may also develop skills in reality testing that have application in other areas of professional and even personal life.

This culture of evaluation that we evaluators take for granted can be quite alien to many of the folks with whom we work. Like people living daily inside any culture, our way of thinking, shaped by the research culture, seems natural and easy to us. However, to practitioners, decision makers, and policymakers, our logic can be hard to grasp and quite unnatural. I'm talking about what appear to be very simple notions that have profound effects on how one views the world. Thinking in terms of what's clear, specific, concrete, and observable does not come easily to people who thrive on, even depend on, vagueness, generalities, and untested beliefs as the basis for action. They're in the majority. Practitioners of evaluation logic are a small minority. The good news is that our way of thinking, once experienced, is often greatly valued. That's part of what creates demand for our services. Exhibit 5.1 provides examples of evaluation logic and values, ways of thinking that undergird evaluation practice (Mark, Greene, and Shaw 2006a:1–3; Fournier 2005a, 2005b, 1995; Scriven 2005a, 1995; Schwandt 2002:59–74; House 1980).

Process use is distinct from use of the substantive findings in an evaluation report. It's equivalent to the difference between learning how to learn versus learning substantive knowledge about something. Learning how to think evaluatively is learning how to learn and think critically, and those who become involved in an evaluation learn by doing. Facilitating evaluative thinking opens up new possibilities for impact that organizations and funders are coming to value because the capacity to engage in this kind of thinking can have more enduring value than a delimited set of findings. This especially resonates for organizations interested in becoming what has come to be called popularly "learning organizations." Learning to see the world as an evaluator sees it often has a lasting impact on those who participate in an evaluation—an impact that can be greater and last longer than the findings from that same evaluation. Findings have a very short "half-life"—to use a physical science metaphor; they deteriorate very quickly as the world changes rapidly. Specific findings typically have a small window of relevance. In contrast, learning to think and act evaluatively can have an ongoing impact. The experience of being involved in an evaluation, then, for those stakeholders actually involved, can have a lasting impact on how they think, on their openness to reality testing, and on how they view the things they do.

How do I know this? Because that's often what intended users tell me when I follow up the evaluations I conduct. Months after an evaluation, I'll talk with clients (intended users) to get their assessments of whether the evaluation achieved its intended uses and to find out what other impacts may have resulted. They often say some version of the following, a response from an experienced and wise program director:

> We used the findings to make some changes in our intake process and improvements in the treatment program. We reorganized parts of the program and connected them together better. But, you know, the big change is in our staff's attitude. They're paying more attention to participant reactions on a daily basis. Our staff meetings are more outcomes oriented and reflective. Staff exchanges about results are more specific and data based. We're more focused. And the fear of evaluation is gone. Doing the evaluation had a profound impact on our program culture. It really did.

Any evaluation can, and often does, have these kinds of effects. What's different about utilization-focused evaluation is that

EXHIBIT 5.1

Evaluative Thinking

The Logic and Values of Evaluation That Can Have Impact on Participants in Evaluation Processes

The logic and values of evaluation derive from principles of systematic inquiry, logical reasoning, and effective communications. The admonitions below constitute a "logic" in the sense that they represent a particular mode of reasoning viewed as valid within the culture of evaluation. They are values in the sense that they are what evaluators generally ascribe value to and believe contribute to effective action. The guidelines and principles below are meant to be illustrative rather than exhaustive of all possibilities.

Be clear	Be clear about goals and purposes; be clear about what's being evaluated, what data will be collected, what judgments are to be made, how results will be used—indeed, be as clear as possible about everything.
Be intentional	Know what you want to do and why. Plan your work and work your plan. Think through what you're doing. Consider contingencies.
Be accountable	Systematically examine the extent to which your intentions and hopes work out as planned and accomplish what you wanted to accomplish.
Be specific	A favorite evaluation clarifying question: "What exactly do you mean by that?"
Focus and prioritize	You can't do or look at everything. Be purposeful in deciding what's worth doing and knowing. Make decisions and own their consequences.
Be systematic	Create a system that covers all priorities. Carefully document what occurs at every stage of decision making and data collection.
Make assumptions explicit	Determine what can and cannot be subjected to empirical test.
Operationalize program concepts, ideas, and goals	The fundamental evaluation challenge is determining how to measure and observe, quantitatively or qualitatively, what is important. Know and specify, operationally, what success will look like—and what constitutes failure. Reality testing becomes real at this point.
Distinguish inputs and processes from outcomes	Confusing processes with outcomes is common. Evaluative thinking looks at the connections between processes and outcomes, and that means distinguishing them and measuring both.
Draw conclusions	Have data to support allegations of fact; provide empirical support based on data and logical explanations for conclusions. This means a commitment to reality testing in which logic and evidence are valued over strength of belief and intensity of emotions.
Separate data-based statements of fact from interpretations and judgments	Interpretations go beyond the data and must be understood as what they are: interpretations. Judgments involve values about what is desirable or undesirable.
Make criteria and standards for judgments explicit	The logical mandates to be clear and specific apply to making criteria and standards explicit.
Limit generalizations and causal explanations to what data support	Overgeneralizations and overly definitive attributions of causality are epidemic outside the culture of research and evaluation.
Cultural sensitivity and cultural competence	Cultural variations and factors are critical to understanding.

the process of actively involving intended users increases these kinds of evaluation impacts. Furthermore, the possibility and desirability of building an organization's capacity to learn from evaluation processes as well as findings can be made intentional and purposeful. In other words, instead of treating process use as an informal offshoot, explicit and up-front attention to the potential impacts of evaluation logic and processes can increase those impacts and make them a planned purpose for undertaking the evaluation. In that way, the evaluation's overall utility is increased.

Process Use Defined

Process use refers to and is indicated by individual changes in thinking, attitudes, and behavior, and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of 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 from going through the thinking process that the evaluation required." As always, the most convincing evidence that learning has occurred is subsequent translation into action. Exhibit 5.2 provides a complete definition of process use. Process use has come to be recognized as an important contribution to evaluation practice and a significant focus of inquiry for research on utilization (Amo and Cousins 2007; Carden and Earl 2007; Cousins 2007, 2003; Cousins and Whitmore 2007; Fleischer 2007; Harnar and Preskill 2007; King 2007a; Lawrenz, Huffman, and McGinnis 2007; Patton 2007, 1999b, 1998; Podems 2007; Cousins and Shulha 2006; Preskill 2005b; Fetterman 2003; Hofstetter and Alkin 2003; Preskill, Zuckerman, and Matthews 2003; Forss, Rebien, and Carlsson 2002; Morabito 2002; Russ-Eft, Atwood, and Egherman 2002; Kirkhart 2000; Preskill and Caracelli 1997).

An Analogy

Before looking in detail at how evaluation processes can affect users, let me suggest an analogy to clarify the distinction between process use and findings use. I hike the Grand Canyon annually. During the days there, my body hardens and my thoughts soften. I emerge more mellow, peaceful, and centered. It doesn't matter which part of the Canyon I hike: the South Rim or North; whether I descend all the way to the Colorado River or stay on the Tonto to explore a side canyon; whether I push strenuously to cover as much territory as possible or plan a leisurely journey; whether I ascend some interior monument like Mount Huethawali or traverse the Supai platform that runs the length of the Canyon—I return different from when I entered. Not always different in the same way. But different.

Let me suggest that the specifics of place are like the findings of an evaluation report. The different places provide different content. From the rim one can view magnificent vistas. Deep within a side canyon, one can see little and feel completely alone. Much of the Canyon is desert, but rare streams and even rarer waterfalls offer a stark contrast to the ancient, parched rock. Each place offers different content for reflection. The substantive insights one receives may well vary by place, time, and circumstance. But quite beyond those variations is the impact that comes from *the very act of reflection*—regardless of content and place. The impacts of reflection and meditation on one's inner sense of self are, for me, analogous to the impacts of engaging in the processes of evaluation, quite apart from the content of the evaluation's findings. In this same sense, for certain

EXHIBIT 5.2

Process Use Defined

Process use occurs when those involved in the evaluation learn from the evaluation process itself or make program changes based on the evaluation process rather than just the evaluation's findings. Process use, then, includes cognitive, attitudinal, and behavior changes in individuals, and program or organizational changes resulting, either directly or indirectly, from engagement in the evaluation process and learning to think evaluatively (e.g., increased evaluation capacity, integrating evaluation into the program, goals clarification, conceptualizing the program's logic model, setting evaluation priorities, improving outcomes measurement). An example of or evidence for process use is when those involved in the evaluation later say something like this: "The impact on our program came not just from the findings but also from going through the thinking process that the evaluation required." As always, the most convincing evidence that learning has occurred is subsequent translation into action. Process use includes the effects of evaluation procedures and operations, for example, the premise that "what gets measured gets done," so establishing measurements and setting targets affects program operations and management focus. These are uses of evaluation processes that affect programs, different from use of specific findings generated by an evaluation.

Process Use as a Usefulism

Process use is best understood and used as a sensitizing concept, or "usefulism" (Safire 2007). A usefulism is an idea or concept that calls our attention to something, but that something takes its meaning from and must be defined within a particular context, like being "middle aged" or manifesting "wisdom." A sensitizing concept, in the tradition of qualitative research (Patton 2007, 2002a), raises consciousness about a possibility and alerts us to watch out for it within a specific context. That's what the concept of "process use" does. The concept process use says, things are happening to people and changes are taking place in programs and organizations as evaluation takes place, especially when stakeholders are involved in the process. Watch out for those things. Pay attention. Something important may be happening. The process may be producing outcomes quite apart from findings. Think about what's going on. Help the people in the situation pay attention to what's going on, if that seems appropriate and useful. In that way, process use can become a matter of intention (Patton 2007).

For variations in operationalizing process use in research on evaluation, see Amo and Cousins (2007) and Cousins (2007).

capacity-building purposes—staff development, program development, organization development (OD)—it doesn't matter so much what the focus of an evaluation is, or the substance of its findings, some impact will come from engaging thoughtfully and seriously in the processes of reflection.

Valuing Process Use

Data use leads to data valuing.

—Cousins, Goh, and Clark (2006)

In working with intended users, it's important to help them think about the potential and desired impacts of how the evaluation will be conducted. Questions about who will be involved take on a different degree of importance when considering that those most directly involved will not only play a critical role in determining the content of the evaluation, and therefore the focus of findings, but they also will be the people most affected by exposure to evaluation logic and processes. The degree of internal involvement,

engagement, and ownership will affect the nature and degree of impact on the program's culture, as will the capacity of the program or organization to engage in evaluation for learning (Harnar and Preskill 2007; King 2007a, 2002; Taut 2007; Baizerman, Compton, and Stockdill 2005; Compton, Baizerman, and Stockdill 2002). How funders and users of evaluation think about and calculate the costs and benefits of evaluation are also affected. The cost-benefit ratio changes on both sides of the equation when the evaluation produces not only findings but also serves immediate programmatic needs like staff development or participant empowerment.

I differentiate six primary uses of evaluation processes: (1) infusing evaluative thinking into an organization's culture; (2) enhancing shared understandings; (3) supporting and reinforcing the program through intervention-oriented evaluation; (4) instrumentation effects (what gets measured gets done); (5) increasing participants' engagement, sense of ownership, and self-determination (participatory and empowerment evaluation); and (6) program or organizational development. Menu 5.1 summarizes these six types of process use. I'll elaborate each, with examples, then consider the challenges and controversies engendered by using evaluation in these ways.

Variety of Process Uses

Infusing Evaluative Thinking into Organizational Culture: Building Evaluation Capacity

What we think, we become.

—Hindu Prince Gautama Siddharta (563–483 BCE), the founder of Buddhism

When evaluation first emerged as a distinct profession in the 1960s and 1970s, the emphasis was on conducting specific evaluation studies for specific purposes. Correspondingly, research on utilization focused on how the findings from those studies were used. Now evaluation use has evolved to include facilitating *evaluative thinking* and building evaluation capacity. Building the evaluation capacity of an organization to support staff in thinking evaluatively means integrating evaluation into the organization's culture. This goes well beyond a focus on using the results of isolated studies. It takes us into the arena of organizational culture, looking at how decision makers and staff incorporate evaluative thinking into everything they do as part of ongoing attention to mission fulfillment and continuous improvement. In his 2001 presidential address at the American Evaluation Association (AEA) annual conference, James Sanders called integrating evaluation into organizational culture "mainstreaming evaluation."

> Mainstreaming refers to the process of making evaluation an integral part of an organization's everyday operations. Instead of being put aside in the margins of work, evaluation becomes a routine part of the organization's work ethic if it is mainstreamed. It is part of the culture and job responsibilities at all levels of the organization. . . . Mainstreaming depends on evaluation being internalized as a value throughout the organization and on an infrastructure that supports and maintains evaluation. (Sanders 2002:254)

Evaluation Capacity Building

Building evaluation capacity goes beyond conducting specific evaluation studies, though conducting specific evaluations by involving program staff can contribute to the organization's evaluation capacity, a premiere example of process use. Evaluation capacity building involves "working intentionally and continuously to create and sustain overall organizational processes that make quality evaluation and its uses routine" (Baizerman, Compton, and Stockdill 2005:39). "The Art, Craft, and Science of Evaluation Capacity Building" is one of the *New Directions for Evaluation* (Compton, Baizerman, and Stockdill 2002).

MENU 5.1

Process Uses

Uses that derive from engaging in an evaluation process in contrast to using evaluation findings

Uses	*Examples*
Infusing evaluative thinking into the organizational culture	Becoming an authentic learning organization Incorporating evaluative questioning into routine decision making Integrating monitoring and evaluation, and linking both to budget and planning cycles Incentives and rewards for evaluation use Building support for evaluation throughout the organization, ongoing capacity development and training in evaluation
Enhancing shared understandings	Agreeing on the program's model and expected outcomes as a result of evaluation questions and determining evaluation priorities *Evaluability assessment*, in which the program gets ready for evaluation by clarifying goals and the program's logic model Managing staff meetings or the program's plan of work around evaluation issues and explicit outcomes Giving voice to different perspectives and valuing diverse experiences
Supporting and reinforcing the program intervention	Building evaluation into program delivery processes in such a way that desired program outcomes are achieved in part through the effects of data collection Participants monitoring their own progress Specifying and monitoring outcomes as integral to working with program participants

Nowhere are evaluative thinking and mainstreaming evaluation as integral to organizational culture better illustrated than in the Corporate Assessment Framework of the International Development Research Centre (IDRC) headquartered in Ottawa, Canada.

IDRC is a public corporation created by the Parliament of Canada in 1970 to help developing countries use science and technology to find practical, long-term solutions to the social, economic, and environmental problems they face. Support is directed toward developing an "indigenous research capacity" to sustain policies and technologies that developing countries need to build healthier, more equitable, and more prosperous societies. IDRC's mission

MENU 5.1 (Continued)

Uses	*Examples*
Instrumentation effects and reactivity	What gets measured gets done so resources and staff efforts are aligned with performance measures and evaluation priorities Using interview protocols to enhance reflection Data collection processes affect program participants and staff intentionally or unintentionally Participants learn from and are affected by evaluation tests, surveys, and interviews Using the data collection process to enhance organizational communications
Increasing engagement, self-determination, and ownership	Participatory and collaborative evaluation Empowerment evaluation Reflective practice Self-evaluation Building evaluation capacity Learning evaluation by doing evaluation
Program and organizational development	Developmental evaluation (See Chapter 8) Making the organization the unit of analysis and organizational effectiveness the focus Looking at the connections between program effectiveness and organizational effectiveness to enhance understanding and support realignment Evaluability assessment and logic modeling used for program design

NOTE: Menu 4.1 (Chapter 4) presents a corresponding menu, "Uses of Evaluation Findings."

is "Empowerment Through Knowledge" (IDRC 2006).

In 1993, IDRC established an Evaluation Unit to support the conduct of evaluations (Carden and Earl 2007). The unit oversees various evaluation and monitoring functions and produces an annual overview of each year's evaluations, synthesizing significant findings for senior management and the Board of Governors. During a retreat in 2001, IDRC's Senior Management Committee expanded the organization's evaluation commitment to include a framework for mission assessment at the overall corporate level. This involved the systematic collection of performance data regarding IDRC's strategic goals and operating principles. To do this, senior managers had to identify those principles, or fundamental ways of doing business, that were expected to permeate all their work in accomplishing their two overall strategic goals: *Indigenous*

Capacity Building and *Policy and Technology Influence*. They committed the organization to monitoring and evaluating not only results in these two strategic goal areas but also the extent to which they were employing their fundamental operating principles. This is where it gets interesting from our point of view because one of those operating principles was *evaluative thinking*. This was the first organization I had encountered that made infusing evaluative thinking into the organizational culture an explicit dimension for performance measurement. (The five other performance areas were Devolution to the South, Regional Presence, Gender, Canadian Partnerships, and Donor Partnerships.)

IDRC's shift in emphasis is a premiere example of process use. In essence, the senior management committed not only to supporting the conduct and use of specific high-quality evaluation studies and management information system data, they made *evaluative thinking* a fundamental way of doing business, infused throughout the culture. What did they mean? Here's what they committed to monitoring as expressed in their working papers operationalizing evaluative thinking.

> Evaluative thinking permeates our work so that we consciously and constantly reflect on project, program, regional, and corporate experience with a view to implementing improvements based on what is learned. Evaluative thinking is evident in the way we clarify goals and design, and conduct and interpret evaluations throughout the organization.
>
> Within programs, evaluative thinking is demonstrated in the implementation of well-focused programs and in the use of high-quality project and program evaluations that feed into program and project decision making. Time and resources are allocated for reflection on evaluation findings and for documenting use of the findings.
>
> In program management, evaluative thinking is evident in our systematic use of evaluation to inform program design and implementation decisions.
>
> Senior management demonstrates evaluative thinking in its support for the adoption of evaluation processes, in its routine demand for the generation of outcomes-based data, and in the use of this and other feedback to expand management viewpoints and inform decisions.
>
> IDRC's partners in the South share our commitment to learning-based evaluation; they exhibit mastery over their own evaluation activities; they serve with competence as external and internal evaluators; and they support and promote advances in the science and art of performance assessment.
>
> The Board of Governors will request and use outcomes-based data in their governance of the Centre.
>
> Senior management will request and use outcomes-based evaluation information in its decision-making processes; engage in the Corporate Assessment Framework processes to reflect on past experiences, analyze empirical data related to the performance areas, and revise actions appropriately to improve performance; share learnings among themselves and with others in the organization; and foster an organizational environment conducive to learning.
>
> Program staff will implement and use the results of high-quality evaluations to improve project and program performance; establish regular processes of reflection (on both successes and failures) at the program level to share lessons and improve future performance; and support the building of capacity in evaluative thinking among partners.

Resources Branch will develop reporting systems to facilitate reflective processes.

Partners will develop effective evaluation systems directed to their own needs and purposes, collaborate with the Centre on implementing evaluations of relevance to both their and our needs, and have the capacity to operate as evaluators.

Evaluation Thinking in Practice

At the same time that IDRC was making evaluative thinking a priority area for overall organizational assessment, senior management was having to face a concrete reality at the most basic level: Project managers were not completing required end-of-project reports. Indeed, they had accumulated a backlog of hundreds of unfinished project completion reports. A variety of carrot-and-stick efforts to get reports completed had failed. Evaluating these efforts, they found that the reports were viewed as an arduous paperwork requirement with no real utility. Project managers didn't get feedback when they did do reports and, given other workload priorities, there were no incentives to complete a report on work already done. All the energy was going into new initiatives rather than recording the details of yesterday's news.

As a part of rethinking the reporting function in the organization with an emphasis on creating opportunities for shared learning, they conducted an inquiry into how and when learning occurs in projects. Project staff said that the most learning takes place at the start of or during a project's life, while the least learning occurs at the end of a project, and different kinds of learning take place throughout a project's varying stages. Drawing on staff at different levels from throughout the organization, a working group was formed to redesign the project-reporting process.

They developed a three-stage process dubbed the "Rolling Project Completion Report" (rPCR) and changed the format, timing, and information input approach. The new system emphasized

The Importance of Evaluative Thinking

Distinguished philosopher Hannah Arendt was especially attuned to the importance of critical thinking as a foundation of democracy. Having experienced totalitarianism in Nazi Germany, then having fled it, she devoted much of her life to studying totalitarianism and its opposite, democracy. Totalitarianism is built on and sustained by deceit and thought control. To resist efforts by the powerful to deceive and control thinking, Arendt believed that people needed to practice thinking. Toward that end she developed "exercises in political thought" (Arendt 1963). She wrote that "experience in thinking . . . can be won, like all experience in doing something, only through practice, through exercises" (p. 4).

From this point of view, might we consider every evaluation an opportunity for those involved to practice thinking? Every evaluation is an opportunity to engage people in thinking evaluatively. In this regard, we might aspire to have evaluation do what Arendt hoped her exercises in political thought would do, namely, "to gain experience in how to think." Her exercises "do not contain prescriptions on what to think or which truths to hold," but rather on the act and process of thinking itself.

learning rather than paperwork accountability. The new approach was pilot tested on a sample of projects from different units throughout the organization. Early in the life of a project, a junior staff member interviews a project officer to gather data about project design, start-up lessons, and issues that will need attention going forward. In the middle of a project, team leaders interview project officers to capture lessons about implementation and interim outcomes, as well as update work on key issues. After the end of a project, senior managers interview project officers to complete the project reports, identify results, and capture any final learnings. Major learnings are highlighted at an Annual Learning Forum. This new rPCR process replaces the old paperwork requirement with an interview process that has people at different levels in the organization talking to each other, learning about each other's work, and sharing lessons. All those involved went through formal interview training, including senior managers, so people share language and understandings about quality interviewing and what kind of cross-organization learning is being sought. The process is designed so that interview responses are entered into the learning system in real time, as the interview takes place, with subsequent opportunities for project managers to make corrections and append supporting documentation and cross-reference information sources (Carden and Earl 2007).

The project report backlog was completely cleared, and feedback about the process is highly positive. The organization-wide process of involving people in reflection and learning reinforces evaluative thinking as a core operating principle while also meeting accountability demands to get reports done in a timely and meaningful fashion. The capacity of staff to engage in evaluation thinking has been systematically enhanced, including deepening their interviewing skills, pattern recognition capabilities, and data interpretation skills. The attention garnered for projects featured at the Annual Learning Forum and the direct involvement of senior management provides additional incentives to take the process seriously and document both learning and results. The Project Completion Reports, long disdained, became a source of energy and enlightenment, and a manifestation of evaluative thinking infused into the organizational culture. This redesign of IDRC's reporting process illustrates nicely the insight of *Future Shock* author Alvin Toffler (1970) who observed, "The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn."

Using Evaluation to Enhance Shared Understandings

Evaluation both depends on and facilitates clear communications. Shared understandings emerge as evaluation logic pushes the senders of messages to be as specific as possible and challenges listeners to reflect on and feed back to senders what they think they've heard. Shared understandings are especially important with regard to expected results. For example, board members and program staff often have different notions of what an agency or program is supposed to accomplish. The processes of clarifying desired ends and focusing staff efforts on accomplishing those ends by evaluating actual accomplishments ought to be primary board functions, but few boards fulfill these functions effectively (Carver 1997).

I'm often asked to facilitate board or staff retreats to help them apply the logic and discipline of evaluation to formulating the organization's mission and goals. The feedback I get is that the questions I pose as an evaluator (e.g., What *specific* results are you committed to achieving and how would you know if you accomplished them? What would success look like?) are different from what they are asked by nonevaluators. It's not so much that other facilitators don't ask these questions, but they don't ask them with the same seriousness and pursue the answers with the same rigor and intensity. The very process of formulating a mission and goals *so they can be evaluated* will usually have an impact on how people think about what they're trying to accomplish, long before data are actually collected to measure results.

A parallel use of evaluation is to increase shared understandings between program managers and line staff. Managers can work with staff under the guidance of an evaluator to establish a monitoring system to help everyone involved stay focused on desired outcomes. While the data from such a system may ultimately support decision making, in the short run, the impact is to focus staff attention and energy on priority outcomes. The process needs to be facilitated in such a way that staff can speak openly about whether board and administrative expectations are meaningful, realistic, and attainable. In other words, done properly, evaluation facilitates shared commitments to results from top to bottom *and* bottom to top for enhanced communication between staff at different levels of program implementation.

The logic and principles of evaluation also can be useful in negotiations between parties with different perspectives. For example, a major foundation was interested in funding an effort to make schools more racially equitable through a process that would engage the community. The school district expressed great interest in such funding but resisted committing to involving community people in schools in any ways that might undermine building-level autonomy or intrude into personnel evaluations of principals. Over a period of several months, the funder and school officials negotiated the project. The negotiations centered on defining what was meant by "greater racial equity." Was the only criterion closing the gap in test scores between students of different races? Should other criteria, like parental involvement, attendance rates, and graduation rates, be included? The funder and school district eventually agreed to focus the project and evaluation on community-based, school-specific action plans, activities, and outcomes rather than a standardized and prescribed set of district-wide, uniform indicators. Part of the reason was to increase the buy-in of teachers and community people on a school-by-school basis since different schools had quite different racial profiles and varying equity challenges. The design of the entire project was changed and made more focused as a result of these negotiations. Applying the logic of evaluation had a major impact on the project's design before any data collection was done, or before findings and a report were produced. Everyone came out of the negotiations clear about what was to happen in the project and how it would be evaluated.

Inadequate specification of desired results reduces the likelihood of attaining those results. Consider how adding a results orientation changed the Request for Proposals announcement of a major environment-oriented philanthropic foundation. In the initial announcement, the foundation wanted to cast the net wide, so it issued a general invitation:

"We seek grant proposals that will enhance the health of specific ecosystems."

The responses varied greatly with many completely missing the mark in the opinion of the foundation staff. But what was the mark? A great deal of time and effort was wasted by hopeful proposal writers who didn't know what criteria to address, and staff spent a lot of time sifting through proposals that had no hope of being funded. The process created frustration on both sides. After a planning session focused on specifying desired results and explicit evaluation criteria, the second announcement was quite a bit more focused:

"We seek grant proposals that will enhance the health of specific ecosystems." Proposals will be judged on the following criteria:

- Clarity and meaningfulness of ecosystem definition
- Private-public sector cooperation
- Action orientation and likelihood of demonstrable impact
- Incorporation of a prevention orientation
- Regional coordination

This set of criteria eliminates basic research proposals, of which a large number were received from universities in the first round, and makes it clear that those seeking grants must submit as cooperative groups rather than as single individuals or entities, also characteristic of a large number of initial proposals. Subsequent announcements became even more specific when focused on specific action priorities, such as pollution prevention. The staff, with training and facilitation, learned to use evaluation logic to articulate desired results and enhance communications with potential grant applicants.

A different use of evaluation to enhance mutual understanding involves designing the evaluation to "give voice" to the disenfranchised, underprivileged, poor, and others outside the mainstream (Weiss and Greene 1992:145). In the evaluation of a diversity project in the Saint Paul Schools, a major part of the design included capturing and reporting the experiences of people of color. Providing a way for African American, Native American, Chicano-Latino, and Hmong (from Laos) parents to tell their stories to mostly white, corporate funders was an intentional purpose of the design, one approved by those same white corporate funders. Rather than reaching singular conclusions, the final report was a multivocal, multicultural presentation of different experiences with and perceptions of the program's impacts. The medium of the report carried the message that multiple voices needed to be heard and valued as a manifestation of diversity (Stockdill et al. 1992). The findings were used for both formative and summative purposes, but the parents and many of the staff were most interested in using the evaluation processes to make themselves heard by those in power. *Being heard was an end in itself, quite separate from use of the findings.*

Russ-Eft et al. (2002) found that in the initial stages of an environmental evaluation, the discussion about the program among diverse program stakeholders "contributed to the organization by enhancing communications" (p. 27). Wadsworth (1995) has reported that evaluation processes can facilitate interactions between service providers and service users in a way that leads to "connectedness" and "dialogue across difference" (p. 9). Each learns to see the service through the other's eyes. In the process, what began as opposing groups with opposing truths is transformed into "an affinity-based community of inquiry" with shared understandings.

Using evaluation to enhance shared understandings is a relatively traditional use of evaluation logic. Let's turn now to

a different and more controversial use of evaluation processes: intervention-oriented evaluation.

Evaluation as Integral to the Program Intervention

Textbooks on measurement warn that measuring the effects of a treatment (e.g., a social program) should be independent of and separate from the treatment itself. For example, participants who take a pretest may perform better in the program than those who do not take the pretest because the pretest increases awareness, stimulates learning, and/or enhances preparation for program activities (instrumentation effects). To account for such test effects, evaluation researchers in the past have been advised to use experimental designs that permit analysis of differences in performance for those who took the pretest compared with a group that did not take the pretest. Integrating data collection into program implementation would be considered a problem—a form of treatment contamination—under traditional rules of research.

Departing from defining evaluation as application of social science methods opens a different direction in evaluation, one that supports integration of evaluation into program processes. Making data collection integral rather than separate can reinforce and strengthen the program intervention. Such an approach also can be cost-effective and efficient since, when evaluation becomes integral to the program, its costs aren't an add-on. This enhances the sustainability of evaluation because, when it's built in rather than added on, it's not viewed as a temporary effort or luxury that can be easily dispensed with when cuts are necessary.

To illustrate this approach, consider the case of a one-day workshop. A traditional evaluation design, based on standard social science standards of rigor, would typically include a pretest and posttest to assess changes in participants' knowledge, skills, and attitudes. Let's suppose you are participating in such a workshop. As the workshop opens, you are told,

> Before we begin the actual training, we want you to take a pretest. This will provide a baseline for our evaluation so we can find out how much you already know and then measure how much you've learned when you take the posttest.

At the end of the day, you are given the same instrument as a posttest. You are told, "Now the workshop is over, but before you leave, we need to have you take the posttest to complete the evaluation and find out how much you have benefited from the training."

The most rigorous design for high internal validity would include, *in addition* to the pre-post treatment group, (1) a control group that takes the pre- and posttests without experiencing the workshops, (2) a control group that gets the posttest only, and (3) a treatment group that gets the posttest only. All groups, of course, should be randomly selected and assigned, and the administration of the test should be standardized and take place at the same time. Such a design would permit measurement of and control for instrumentation effects.

Let me now pose a contrary example of how the evaluation might be handled, a design that fully integrates the evaluation data collection into the program delivery, that is, a design that makes the data collection part of the workshop rather than separate from

and independent of the workshop. In this scenario, the workshop begins as follows:

> The first part of the workshop involves your completing a self-assessment of your knowledge, skills, and attitudes. This will help you prepare for and get into thinking about the things we will be covering today in your training.

The workshop then proceeds. At the end of the day, the workshop presenter closes as follows:

> Now the final workshop activity is for you to assess what you have learned today. To that end, we are going to have you retake the self-assessment you took this morning. This will serve as a review of today and let you see how much you've learned.

In this second scenario, the word *evaluation* is never mentioned. The pre- and post assessments are explicitly and intentionally part of the workshop in accordance with adult learning principles (Brookfield 1990; Knox 1987; Schön 1987; Knowles et al. 1985). We know, for example, that when participants are told what they will learn, they become prepared for the learning; learning is further enhanced when it is reinforced both immediately and over the long term. In the second scenario, the self-assessment instrument serves *both* the function of preparing people for learning and as baseline data. The posttest serves the dual functions of learning reinforcement and evaluation. Likewise, a 6-month follow-up to assess retention can serve the dual functions of learning reinforcement and longitudinal evaluation.

The methodological specialist will note that the second scenario is fraught with threats to validity. However, the purpose of data collection in this second scenario is not only assessment of the extent to which change has occurred, but increasing the likelihood that change will occur. It does not matter *to these particular intended users (the workshop instructors)* how much of the measured change is due to pretest sensitization versus actual learning activities, or both, as long as the instrument items are valid indicators of desired outcomes. Moreover, in the second scenario, the data collection is so well integrated into the program that there are no separate evaluation costs except for the data analysis itself. Under the second scenario, the administration of the pretest and posttest is a part of the program such that *even if the data were not analyzed for evaluation purposes, the data collection would still take place,* making evaluation data collection highly cost-effective.

Principles of Intervention-Oriented Evaluation

I have called this process *intervention-oriented evaluation* to make explicit the direct and integral connection between data collection and program results. A program is an intervention in the sense that it is aimed at changing something. The evaluation becomes part of the programmatic intervention to the extent that the way it is conducted supports and reinforces accomplishing desired program goals.

The primary principle of *intervention-oriented evaluation* is to build a program delivery model that logically and meaningfully interjects data collection in ways that enhance achievement of program outcomes, while also meeting evaluation information needs. We followed this principle in evaluating a wilderness program that aimed to transform traditional college administrators into leaders in *experiential education*. Participants were university presidents, college deans, and department heads with no previous wilderness experience. They hiked 10 days in the Gila Wilderness of New Mexico in the fall, climbed the Kofa Mountains of Arizona in

the winter, and rafted the San Juan River in Utah in the spring. During these trips, participants kept journals for reflection. The program's philosophy was, "One doesn't just learn from experience; one learns from *reflection* on experience." The process of journaling was part of the program intervention, but also a prime source of qualitative evaluation data capturing how participants reacted to and were changed by project participation. In addition, participants were paired together to interview each other before, during, and after each wilderness experience. These interviews were part of the project's reflection process, but also a source of case data for evaluation. The evaluation process thus became part of the intervention in providing participants with experiences in *reflective practice* (Schön 1987, 1983). Indeed, it was on this project that I first learned how profoundly in-depth interviews can affect people. Such personal, intensive, and reflective data collection is an intervention. In intervention-oriented evaluation, such data collection is designed to reinforce and strengthen the program's impact.

Another quite different example comes from an intervention-designed evaluation of an international development effort called the Caribbean Agricultural Extension Project, funded by the U.S. Agency for International Development. The project aimed to improve national agricultural extension services in eight Caribbean countries. The project began with a rapid reconnaissance survey to identify the farming systems in each participating island. This involved an interdisciplinary team of agricultural researchers, social scientists, and extension staff doing fieldwork and interviewing farmers for a period of 10 days to identify extension priorities for a specific agro-ecological zone. This process served as the basis for needs assessment and program development. It was also, quite explicitly and intentionally, an intervention in and of itself in that the process garnered attention from both farmers and agricultural officials, thereby beginning the extension mobilization process. In addition, the rapid reconnaissance survey served the critical evaluation function of establishing baseline data. Subsequent data on the effects of extension and agricultural development in the zone were compared against this baseline for evaluation purposes. Yet it would have been much too expensive to undertake this kind of intensive team fieldwork simply for purposes of evaluation. Such data collection was practical and cost-effective because it was fully integrated into other critical program processes.

Once the various farming systems were identified and the needs of farmers had been specified within those systems, the extension staff began working with individual farmers to assess their specific production goals. This process included gathering data about the farmer's agricultural enterprises and household income flows. With these data in hand, extension agents worked with farmers to set realistic goals for change and to help farmers monitor the effects of recommended interventions. The program purpose of using this approach, called a *farm management approach*, was to individualize the work of extension agents with farmers so that the agent's recommendations were solidly grounded in knowledge of the farm and household situation, including labor availability, land availability, income goals, and past agricultural experiences. These data were necessary for the extension agent to do a good job of advising farm families about increasing their productivity.

These same data were the baseline for measuring the program's impact on individual farmers for evaluation purposes.

The collection of such data for farm management purposes required training of agents and a great deal of time and effort. It would have been enormously expensive to collect such data independently, solely for purposes of evaluation. However, by establishing a record-keeping system for individual farmers that served a primary extension purpose, the project also established a record-keeping system for evaluation purposes. By aggregating the data from individual households, it was possible to analyze system-level impact over time. The data aggregation and comparative analysis were above and beyond the main program purpose of collecting the data. However, without that program purpose, the data would have been much too expensive to collect solely for evaluation of the system.

The program staff also used the evaluation design formulated by the external evaluators as the framework for their plan of work, which set the agenda for monthly staff meetings and quarterly staff reports (an example of using evaluation to enhance and focus communications). In this way, the evaluation priorities were kept before the staff at all times. As a result, the evaluation process improved program implementation from the very beginning by focusing staff implementation efforts.

Still another powerful example of intervention-oriented evaluation comes from the Hazelden Foundation, a chemical dependency treatment program in Minnesota. Part of the program intervention includes helping clients and their significant others identify their chemical abuse patterns. A self-assessment instrument serves this purpose while also providing baseline data on chemical use. After residency treatment, all clients and significant others receive follow-up surveys at 6 months, 1 year, and 2 years. The follow-up surveys not only provide outcome data on program effectiveness but they also remind clients and their significant others to assess their current chemical use behaviors. Clients who have relapsed into drugs or alcohol abuse are invited to contact Hazelden for support, assessment, and possible reentry into treatment. Thus, the follow-up survey is a mechanism for reinforcing treatment and extending an offer of new help. Many clients respond to this contact and seek additional help. For that reason, the survey is sent to all former clients, not just the small random sample that would be sufficient if the survey provided only evaluation data.

In my experience, program funders, managers, and staff can become very excited about the creative possibilities for integrating evaluation into a program in such a way that it supports and reinforces the program intervention. Not only does this make the evaluation process more useful, but it also often makes the evaluation findings more relevant, meaningful, accessible, and useful. Yet this approach can be controversial because the evaluation's credibility may be undercut by concerns about whether the data are sufficiently independent of the treatment to be meaningful and trustworthy; the evaluator's independence may be suspect when the relations with staff and/or participants become quite close; and the capacity to render an independent, summative judgment may be diminished. These are considerations to discuss with intended users and evaluation funders in deciding the relative priority of different potential uses of evaluation and in reviewing the principles of intervention-oriented evaluation (see Exhibit 5.3). It is also helpful to examine data or other information resources that already exist within a program that can be used in a different way in support of the program.

EXHIBIT 5.3

Principles of Intervention-Oriented Evaluation

- The evaluation is designed to support, reinforce, and enhance attainment of desired program outcomes.
- Evaluation data collection and use are integrated into program delivery and management. Rather than being separate from and independent of program processes, the evaluation is an integral part of those processes.
- Program staff and participants know what is being evaluated and know the criteria for judging success.
- Feedback of evaluation findings is used to increase individual participant goal attainment as well as overall program goal attainment.
- There are no or only incidental add-on costs for data collection because data collection is part of program design, delivery, and implementation.
- Evaluation data collection, feedback, and use are part of the program model, that is, evaluation is a component of the intervention.

Instrumentation Effects and Reactivity

Intervention-oriented evaluation as a form of intended process use occurs when evaluation data collection is *intentionally* made integral to the program treatment as in the Hazelden follow-up assessments described above. Process use can also be unintentional, especially where those participating in evaluation data collection react to completing tests or filling out evaluation questionnaires or are affected by interviews. Instrumentation effects, or reactivity, call attention to the ways in which people are affected by taking tests, completing surveys, or being interviewed.

In-depth, open-ended interviews, for example, can have a powerful effect on people (Patton 2002a:chap. 7). An intensive, probing interview that invites a program participant or staff member to reflect on and share experiences and perspectives can affect him or her in unexpected and unpredictable ways. Emotions may be aroused, assumptions examined, and changes in behavior stimulated. A good interview often opens up or brings to the surface thoughts, feelings, knowledge, and experience, not only to the evaluator conducting the interview but also to the person being interviewed. The process of being taken through a directed, reflective process can leave interviewees realizing things about themselves that they were not fully aware of before the interview. Two hours or more of thoughtfully reflecting on a program experience can evoke a lot of thoughts and feelings, even stimulate commitment to change some behavior pattern. Such a change would be *an instrumentation effect*—an effect of the interview process and experience.

The purpose of an evaluation interview is first and foremost to gather data, not change people. Skilled interviewers take a nonjudgmental stance and communicate neutrality so that the person being interviewed feels comfortable saying what they really think and feel. The evaluator's job is

to judge the program not the program participants who provide evaluative reflections. Neither is an evaluation interviewer a therapist. Staying focused on the purpose of the interview is critical to gathering high-quality data. Still, it is common for program participants being interviewed to ask for advice, approval, or confirmation. This is a clue that the evaluation process—both the interview and, perhaps, the evaluator—is stimulating a reaction.

While the evaluation interviewer strives to present a stance of neutrality, that same interviewer is also attempting to establish rapport and show empathy, that is, understanding as opposed to sympathy. In so doing, the evaluation interviewer is not a cold slab of granite—unresponsive to revelations of great suffering, fear, and pain that may unfold during an interview. Let me take you inside this dilemma. In a major farming needs assessment project to develop agricultural extension programs for distressed farm families during the Midwestern farm crisis of the mid-1980s, I was part of a team of 10 interviewers (working in male-female pairs) who interviewed 50 farming couples. Many of these couples were struggling economically and emotionally. They were losing their farms. Their children had left for the city. Their marriages were under stress. The 2-hour interviews traced their family history, their farm situation, their community relationships, and their hopes for the future. Sometimes questions would lead to husband-wife conflict. The interviews would open old wounds, lead to second-guessing decisions made long ago, or bring forth painful memories of dreams never fulfilled. People reacted to the reflective process of the interview. Because we were university sponsored and clearly well educated, people often asked for advice—what to do about their finances, their children, government subsidy programs, even their marriages. But we were not there to give advice. Our task was to get information about needs that might, or might not, lead to new programs of assistance. Could we do more than just ask our questions and leave? Yet as researchers, could we justify in any way intervening? Yet again, our interviews were already an intervention, stimulating people to think about past choices and future options. Such are the ethical dilemmas that can derive from the power of evaluation data collection, especially in-depth interviewing.

What we decided to do was leave each family a packet of information about resources and programs of assistance, everything from agricultural referrals to financial and family counseling. To avoid having to decide which couples really needed such assistance, we left the information with all couples—separate and identical packages for both husband and wife. When interviewees asked for advice during the interview, we could tell them that we would leave them referral information at the end of the interview. This was also a way of offering them something for giving us their time in participating in the needs assessment.

While interviews may be intrusive in reopening old wounds, they can also be healing. In doing follow-up interviews with families who had experienced child sexual abuse, we found that most mothers appreciated the opportunity to tell their stories, vent their rage against the system, and share their feelings with a neutral, but interested, listener. Our interviews with elderly residents participating in a program to help them stay in their homes and avoid nursing home institutionalization typically lasted much longer than planned because the elderly interviewees longed to have company and talk. When interviewees are open and willing to talk, the power of interviewing poses new risks. People will tell you things they never intended to tell

you. This can be true even with reluctant or hostile interviewees, a fact depended on by journalists. Indeed, it seems at times that the very thing someone is determined *not* to say is the first thing they tell, just to release the psychological pressure of secrecy or deceit. Interviews can become confessions, particularly under the promise of confidentiality. All these are *reactions to the process of being interviewed.*

These reactions can affect the participants' understanding of a program. People being interviewed learn things from the questions asked. If you ask a mother in an early childhood education program whether she is engaging with other mothers outside the formal sessions, she may respond, as one quite shy mother once did to me, "No, but that's a good idea. I'd like to do that." Then, looking quizzical, "Are we supposed to be doing that? I get it. That's why we all exchanged phone numbers." She was learning about the program from the evaluation interview.

This is a qualitative example of a classic concern about pre-post knowledge tests in which participants learn from taking the test, an issue raised earlier.

Learning from being interviewed or taking a test is an instrumentation effect. Questionnaires can also affect people, solidifying attitudes about a program or stimulating discussions among those who complete the questionnaires: "How did you answer that question about how the program can be improved?" Such exchanges of views can affect how a group of participants engage with staff going forward.

The performance measurement mantra, "What gets measured gets done," is a recognition of process use and measurement reactivity. When a teacher announces a test and says, "Here's what will be on the test and here's what I'll be looking for," that teacher is manifesting the performance measurement principle that *what gets measured gets done*. Some weight loss programs have participants set goals and then weigh in weekly in front of other participants as a way of increasing commitment to losing weight. Weighing oneself is a measurement. It is also feedback about progress aimed at encouraging goal achievement. Weighing oneself in front of other participants increases the likelihood of reacting to the measurement process. That's why it is done, to harness the desire to avoid shame and embarrassment in service of goal attainment. Such regular monitoring becomes part of the intervention.

In program performance monitoring, what gets measured gets done means that program resources and staff efforts are meant to be focused on moving the indicators. Staff, therefore, is expected to react to what is being measured. In the chapter on goals-based evaluation, we'll look in greater depth at the challenges of choosing and using performance indicators, including concerns about goal displacement (doing what is measurable rather than what's important) and corruption of indicators. The point here is that process use occurs when the measurement process itself affects what people do and when the evaluation data collection instruments—tests, questionnaires, interviews, journals—affect how participants experience a program and what outcomes they attain.

The interview-based project completion report process developed by IDRC, described earlier in this chapter, exemplifies intentional instrumentation effects. The IDRC staff being interviewed are supposed to become reflective as a result of the interviews. Indeed, the interviews are designed to increase reflectivity—a form of intentional and desired instrument reactivity. The IDRC staff conducting the interviews are also supposed to be affected in that they are learning from the people they are interviewing. The data collection

What Gets Measured Gets Done—For Better or Worse

Examples of "what gets measured gets done"—if you measure the wrong thing, you do the wrong thing:

If a highway department measures the number of potholes filled as a performance indicator for street repair it may reward quickly done, temporary pothole filling to show "good numbers" versus the more time-intensive and time-consuming repairs that result in a longer-term fix.

If an office manager insists that front line staff answer every call within three rings, the results may be a lot of answered calls where most of the callers are put on hold for excessive amounts of time. Answering the phone and actually responding to a call involve different outcomes and different measures.

"We found that fewer measures, but measures that were developed in collaboration with the people doing the work being measured, were more successful."

Theresa N. Westover, Ph.D.
California Department of Education
Policy & Evaluation Division
EvalTalk listserv posting, July 31, 2006

process is designed to enhance communications and learning throughout IDRC, including across divisions and up and down the hierarchy of staff responsibility. While the interview content generates findings about projects, the interview *process* enhances communications and facilitates organizational learning. Thus, the IDRC project completion report design includes both findings use and process use.

Now, let's examine the intentional use of entire evaluation processes, not just data gathering, to engage participants more fully, empower them, and build their capacity to think evaluatively.

Supporting Engagement, Self-Determination, and Ownership: Participatory, Collaborative, and Empowerment Evaluation

Early in my career, I was commissioned by a Provincial Deputy Minister in Canada to undertake an evaluation in a school division he considered mediocre. I asked what he wanted the evaluation to focus on. "I don't care what the focus is," he replied.

> I just want to get the people engaged in some way. Education has no life there. Parents aren't involved. Teachers are just putting in time. Administrators aren't leading. Kids are bored. I'm hoping evaluation can stir things up and get people involved again.

That's how the evaluation of the Frontier School Division, described in Chapter 2, began.

The processes of participation and collaboration have an impact on those who participate beyond whatever tasks they may accomplish by working together. In the *process* of participating in an evaluation, participants are exposed to and have the opportunity to learn the logic of evaluation and the discipline of evaluation reasoning. Skills are acquired in problem identification, criteria specification, and data collection, analysis, and interpretation. Acquisition of evaluation skills and ways of thinking can have a

What Gets Measured Gets Measured

"What gets measured gets done."
Management speak. Baloney. Nonsense.
Hopeful, to be sure. Visionary, perhaps.
Desirable, maybe. Consultant's promise, surely.
Nice PowerPoint slide.
Also a lie.
The causal link is problematic at best.
What gets measured gets measured.
What gets done gets done.
What gets measured may get done
and what gets done may get measured.
But don't confuse measuring with doing,
or doing with measuring.
Measuring something does not assure it gets done.
any more than
counting how many times you cast a fishing line
assures that you will catch a fish.

—Halcolm

longer-term impact than the use of findings from a particular evaluation study.

Moreover, people who participate in creating something tend to feel more ownership of what they have created, make more use of it, and take better care of it. Active participants in evaluation, therefore, are not only more likely to feel ownership of their evaluation findings but also of the evaluation process itself. Properly, sensitively, and authentically done, it becomes *their* process.

Participants and collaborators can be staff and/or program participants (e.g., clients, students, community members). Sometimes administrators, funders, and others also participate, but the usual connotation is that the primary participants are "lower down" in the hierarchy. In that sense, participatory evaluation is bottom-up.

In 1995, evaluators interested in "Collaborative, Participatory, and Empowerment Evaluation" formed a Topical Interest Group within the AEA. What these approaches have in common is a style of evaluation in which the evaluator becomes a facilitator, collaborator, and teacher in support of program participants and staff engaging in their own evaluation. While the findings from such a participatory process are intended to be used for program improvement, the more immediate impact is to use the evaluation process to increase participants' sense of being in control of, deliberative about, and reflective on their own lives and situations.

The labels "participatory evaluation" and "collaborative evaluation" mean different things to different evaluators. Some use these

phrases interchangeably or as mutually reinforcing concepts (e.g., Dugan 1996; Powell et al. 1989; Whitmore and Kerans 1988). Wadsworth (1993) distinguishes "research on people, for people, or with people" (p. 1). Whitmore (1988) has defined the participatory approach as combining "social investigation, education, and action with the ultimate purpose of engendering broad community and social change" (p. 3). Whitmore worked with a community-based team and contended that, through the evaluation process, participants gained not only new knowledge and skills but also created a support network among themselves and gained a greater sense of self-efficacy. In a seminal article, Cousins and Whitmore (2007) distinguish three separate dimensions of participation and collaboration: (1) who controls the evaluation process (a researcher-practitioner continuum), (2) stakeholder selection for participation (a continuum from all legitimate groups to just primary intended users), and (3) depth of participation (a continuum from consultation to deep participation).

In the mid-1980s, several international grassroots development organizations advocated participatory evaluation as a tool for community and local leadership development in addition to being a management tool (PACT 1986). In advocating for participatory evaluation, the Evaluation Sourcebook of the American Council of Voluntary Agencies for Foreign Service asserted, "Participation is what development is about: gaining skills for self-reliance" (ACVAFS 1983:12). Thus, in developing countries, participatory monitoring and evaluation (M&E) has been linked to community development and capacity building (e.g., Salmen and Kane 2006; PREVAL 2006; Wageningen International UR 2006; Kuzmin 2005; Vernooy, Qui, and Jianchu 2003). More industrialized countries, where notions of "value-free" social science have long been dominant, have come to this idea of linking evaluation participation with empowerment more slowly, and, as we shall see later, the notion remains controversial.

Norman Uphoff (1991) published *A Field Guide for Participatory Self-Evaluation* aimed at grassroots community development projects. After reviewing a number of such efforts, he concluded,

> If the process of self-evaluation is carried out regularly and openly, with all group members participating, the answers they arrive at are in themselves not so important

The Global Movement toward More Participatory Approaches

As the World Bank task team leader starts her long flight from Washington to Accra, perhaps she reflects on how much the concept of development has changed over the past 25 years, and what the changes mean for how she does her work. It has been a long time since anyone in development could work only with governments and technical experts to create and implement plans for a country or sector. Now, she knows that social and cultural factors often carry as much or more weight than economic ones. Local stakeholders' concerns and insights are valued, indeed are often critical, in planning and carrying out development programs. Her work and that of her country government partners must include those stakeholders' views now that international development agencies, governments, and civil society organizations are clear that stakeholders have rights and roles in development, and that their involvement is critical, both technically and morally.

SOURCE: Lawrence Salmen and Eileen Kane (2006:7). *Bridging Diversity: Participatory Learning for Responsive Development.*

> as what is learned from the discussion and from the process of reaching consensus on what questions should be used to evaluate group performance and capacity, and on what answers best describe their group's present status. (P. 272)

Here is clear support for the central premise of this chapter: *The process of engaging in evaluation can have as much or more impact than the findings generated.* It was not a group's specific questions or answers that Uphoff found most affected the groups he observed. It was the process of reaching consensus about questions and engaging with each other in the meaning of the answers turned up. The process of participatory self-evaluation, in and of itself, provided useful learning experiences for participants.

Since no definitive definitions exist for "participatory" and "collaborative" evaluation, these phrases must be defined and given meaning in each setting where they're used. Exhibit 5.4 presents what I consider the primary principles of participatory evaluation. This list can be a starting point for working with intended participants to decide what principles they want to adopt for their own process.

Cousins and Earl (1995) examined how participatory and collaborative approaches contribute to increased use of findings: "Unlike emancipatory forms of action research, the rationale for participatory evaluation resides not in its ability to ensure social justice or to somehow even the societal playing field but in the utilization of systematically collected and socially constructed knowledge" (p. 10). They then ventured beyond increased use of findings when they discussed how participation helps create a learning organization. Viewing participatory

EXHIBIT 5.4

Principles of Participatory Evaluation

- The evaluation process involves participants in learning evaluation logic and skills, for example, goal setting, establishing priorities, focusing questions, interpreting data, data-based decision making, and connecting processes to outcomes.
- Participants in the process *own* the evaluation. They make the major focus and design decisions. They draw and apply conclusions. Participation is real, not token.
- Participants focus the evaluation on process and outcomes they consider important and to which they are committed.
- Participants work together as a group, and the evaluation facilitator supports group cohesion and collective inquiry.
- All aspects of the evaluation, including the data, are understandable and meaningful to participants.
- Internal, self-accountability is highly valued. The evaluation, therefore, supports participants' accountability to themselves and their community first, and external accountability secondarily, if at all.
- The evaluator is a facilitator, collaborator, and learning resource; participants are decision makers and evaluators.
- The evaluation facilitator recognizes and values participants' perspectives and expertise and works to help participants recognize and value their own and each other's expertise.
- Status differences between the evaluation facilitator and participants are minimized.

evaluation as a means of creating an organizational culture committed to ongoing learning has become an important theme in literature linking evaluation to learning organizations and capacity building (e.g., Harnar and Preskill 2007; King 2007a, 2007b, 2002, 1995; PREVAL 2006; Fetterman and Wandersman 2005; Kuzmin 2005; Owen 2005; Podems 2005; Preskill 2005a, 2005b; Preskill and Russ-Eft 2005; Cousins 2004; Baker and Sabo 2004; Preskill and Torres 1999; Sonnichsen 1993; Leeuw et al. 1993; Aubel 1993). "The goal of a participatory evaluator is eventually to put him or herself out of work when the research capacity of the organization is self-sustaining" (King 1995:89). Indeed, "the self-evaluating organization" (Wildavsky 1985) constitutes an important direction in the institutionalization of evaluation logic and processes.

Cousins (2001) studied and reported perspectives on collaborative evaluation from 67 pairings of North American evaluators and nonevaluator program practitioners who had participated together on a specific evaluation project. The results highlighted

> the benefits of evaluation process as a distinct source of impact on programs and, in particular, stakeholders associated with them. . . . Emerging evidence suggests that the logic of evaluation and systematic inquiry can be integrated into program, organizational, and community culture, but that such eventualities are most likely to occur within sustained evaluation activities that involve evaluators working in collaboration with non-evaluator stakeholder participants." (Pp. 127–28)

And just what does Cousins (2001) mean by "sustained evaluation activities" that involve collaboration?

> In our approach to collaborative evaluation (called practical participatory evaluation) primary users of evaluation data participate directly in the evaluation process from start to finish, including many of the technical activities such as instrument development, data collection, processing, and interpretation and reporting. We suggest that engagement in such activities engenders deep levels of understanding, by evaluators and program practitioners alike. . . . Collaborative evaluation of this sort is consistent with a utilization-oriented, problem-solving approach. (Pp. 115–16)

Participatory evaluation involves a partnership between the evaluator and those who participate in the evaluation (McGarvey 2007; Baker and Sabo 2004). These partnerships can be at the organizational level or with entire communities, as in a community development process (Cabaj 2007; Ridde 2006). As with any partnerships in any arena, building and maintaining a partnership is not easy, despite the best of intentions. Participatory evaluation partnerships can be particularly challenging in part because of underlying fears, bad past experiences with evaluation, resistance to reality testing, and cultural norms that undercut openness and questioning (Podems 2005). Facilitating participatory processes adds layers of complexity to the already complex tasks of evaluation. Nor do all evaluators have the skills and temperament to successfully engage in and facilitate a participatory evaluation. Some evaluators can't imagine anything more horrific than spending time with a group of untrained, nonresearch-oriented laypeople designing an evaluation. "Why would I want to do that?" A colleague once asked me after a panel on the topic, his face filled with disdain. "I can't imagine a bigger waste of time." Such evaluators resonate with the sentiment of Canadian piano virtuoso Glenn Gould who suffered from such horrible stage fright that he abandoned public performing completely.

"To me," he said, "the ideal artist-to-audience relationship is one to zero" (Lahr 2006:38). Many evaluators feel similarly about the desired ratio between the evaluator and those participating in the evaluation process: The ideal evaluator-to-participant relationship is one to zero.

Utilization-focused evaluation is inherently participatory and collaborative in actively involving primary intended users in all aspects of the evaluation. Evidence presented in earlier chapters has demonstrated the effectiveness of this strategy for increasing use of findings. The added emphasis of this chapter is how participation and collaboration can lead to an ongoing, longer-term commitment to using evaluation logic and building a culture of learning in a program or organization. Making this kind of process use explicit enlarges the menu of potential evaluation uses. How important this use of evaluation should be in any given evaluation is a matter for negotiation with intended users. The practical implication of an explicit emphasis on creating a learning culture as part of the process will mean building into the evaluation attention to and training in evaluation logic and skills.

When designing and negotiating participatory evaluations, one can use the three dimensions discussed earlier (Cousins 2003) that define important variations in the degree of participation and collaboration between evaluators and nonevaluators: (1) the control of the process—how much is practitioner controlled versus researcher controlled; (2) the scope of participation—many diverse stakeholders versus select primary intended users; and (3) depth of participation—great and deep participation by nonevaluator practitioners versus a more consultative, advice-giving, and perspective-sharing role. There are not right places to be on these three dimensions. Rather, the depth and nature of participation depend on the situation, the purpose of the evaluation, the skills of those involved, and the degree to which participant learning is a major intended outcome of the evaluation. Suzanne Callahan (2005) has provided excellent "case studies in the art of evaluation" that take us inside participatory evaluation in real-world practice.

Not all references to "participatory" or "collaborative" evaluation make the link to participant learning—and the labels are sometimes misused. I have seen evaluations described as "participatory" simply because the evaluator surveyed program participants to get feedback about the program. That does *not* qualify as participatory evaluation, which requires some degree of interaction with and involvement of participants in the evaluation design itself, not just filling out the evaluator's survey questions or responding to interviews.

Levin (1993) distinguished three purposes for collaborative research: (1) the pragmatic purpose of increasing use, (2) the philosophical or methodological purpose of grounding data in practitioner's perspectives, and (3) the political purpose of mobilizing for social action. A fourth purpose, identified here, is teaching evaluation logic and skills, or more generally, *building evaluative capacity.* Cousins et al. (2004) found an empirical relationship between evaluation capacity and organizational learning in schools. Preskill and Russ-Eft (2005) have identified and described 72 activities that can be used in *Building Evaluation Capacity.* Empowerment evaluation makes capacity building a priority of collaboration.

Empowerment Evaluation

The theme of the 1993 AEA national conference was "Empowerment Evaluation."

Factors Affecting Learning from an Evaluation

The theme chosen by President Hallie Preskill for the 2007 Annual Conference of the American Evaluation Association was "Learning to evaluate . . . evaluating to learn."

Several factors appear to influence the likelihood that those involved in evaluation processes will learn from their participation. These include factors related to the following:

1. *How evaluation meetings are facilitated.* This involves the intentionality of learning from the evaluation process, the amount and quality of dialogue and reflection, the meeting facilitators' group process skills, the degree of trust among participants, and how much time is given to discussing various issues.
2. *The extent to which, and the ways in which, management and leadership support participants' involvement in the evaluation process*. This involves expectations managers have for participants to share their learning with others in the organization or community, and how they are rewarded for sharing and using what they have learned.
3. *Participants' personal characteristics and experiences with evaluation in the program being evaluated*. These include participants' motivation to engage in the evaluation process, their position, their rank, their previous training in evaluation, and the belief that evaluation findings will be used.
4. *The frequency, methods, and quality of communications between and among stakeholder participants*.
5. *Organizational characteristics*. These include the extant degree of organizational stability, external demands, constraints, and threats in the extent to which the organization supports evaluation work.

If process use is supported, nurtured, and studied, it may lead not only to individual learning but also to team and organizational learning.

SOURCES: Preskill (2005b: 328, 2007).

David Fetterman (1993), AEA President that year, defined empowerment evaluation as "the use of evaluation concepts and techniques to foster self-determination. The focus is on helping people help themselves" (p. 115).

> Self-determination, defined as the ability to chart one's own course in life, forms the theoretical foundation of empowerment evaluation. It consists of numerous interconnected capabilities that logically follow each other . . . : the ability to identify and express needs, establish goals or expectations and a plan of action to achieve them, identify resources, make rational choices from various alternative courses of action, take appropriate steps to pursue objectives, evaluate short- and long-term results (including reassessing plans and expectations and taking necessary detours), and persist in pursuit of those goals. (Fetterman 1994a:2)

These skills are used to realize the group's own political goals; through self-assessment and a group's knowledge of itself it achieves accountability unto itself as well as to others (Fetterman, Kaftarian, and Wandersman 1996; Fetterman 1994b). With experience and reflection, empowerment evaluation has come to focus on 10 principles (Fetterman and Wandersman 2005):

1. Improvement
2. Community ownership
3. Inclusion
4. Democratic participation
5. Social justice
6. Community knowledge
7. Evidence-based strategies
8. Capacity building
9. Organizational learning
10. Accountability

Fetterman (2005a) explains that there is no absolute ranking or ordering of these principles. "However, there is a logical flow of the principles in practice" (p. 4). The principles are interdependent.

> As a general rule, the quality [of an empowerment evaluation] increases as the number of principles are applied, because they are synergistic. Ideally each of the principles should be enforced at some level. However, specific principles will be more dominant than others in each empowerment evaluation. The principles that dominate will be related to the local context and purpose of evaluation. Not all principles will be adopted equally at any given time or for any given project. (P. 9)

One of the conceptual developments in empowerment evaluation has been recognizing its impacts on individuals, organizations, and communities as a form of process use. As an example, in one evaluation Fetterman (2005b) reports the following process uses:

> A culture of evidence evolved from the discussions and self-assessments. The community of learners engaged in a dialogue about the status of their efforts. This engagement helped them build and refine skills in discourse and reasoning. The skills were generalizable to many other facets of their life. It also contributed to building a trusting relationship in the process. (P. 102)

Empowerment evaluation is most appropriate where the goals of the program include helping participants become more self-sufficient and personally effective. In such instances, empowerment evaluation is also intervention-oriented in that the evaluation is designed and implemented to support and enhance the program's desired outcomes. Weiss and Greene (1992) have shown how "empowerment partnerships" between evaluators and program staff were particularly appropriate in the family support movement because that movement emphasized participant and community empowerment. Orthner et al. (2006) conducted a cross-national experimental design comparison and found that organizational learning, staff empowerment, and program outcomes were related, and that it was especially important to focus the learning on program outcomes for, in this case, improved lives of children in the program.

As with the label "participatory evaluation," merely labeling an evaluation an empowerment evaluation does not make it empowering. Miller and Campbell (2006) studied 47 published studies labeled empowerment evaluations. They found wide variation in what was done and which principles were followed, including a case of an evaluation that was designed and executed solely by an evaluator

> with no input or involvement from stakeholders. . . . In this particular case, the evaluator indicated that the project was an empowerment evaluation because by allowing a disenfranchised population to respond to a survey, the population was afforded a voice. (P. 306)

I facilitated a cluster team evaluation of 34 programs serving families in poverty (Patton 1993). A common and important outcome of those programs was "increased intentionality"—having participants end up with a plan, a sense of direction, taking responsibility for their lives, and a commitment to making progress. Increased intentionality began with small first steps. Families in poverty often feel stuck where they are or are experiencing a downward spiral of worsening conditions and ever-greater hopelessness. These programs commonly reported that it was a major achievement to give people a sense of hope manifest in a concrete plan that participants had developed, understood, and believed they could accomplish. "Increased intentionality" is a commitment to change for the better and a belief that such a change is possible. Thus, the programs collectively placed a great deal of emphasis on developing such skills as goal setting, learning to map out strategies for attaining goals, and monitoring progress in attaining personal goals. The programs' evaluations were built around these family plans and supported them. Developing family plans was not an end in itself, but the ability and willingness to work on a plan emerged as a leading indicator of the likelihood of success in achieving longer-term outcomes. Creating and taking ownership of a plan became milestones of progress. The next milestone was putting the plan into action.

Another empowering outcome of participatory evaluation is forming effective groups for collective action and reflection. For example, social isolation is a common characteristic of families in poverty. Isolation breeds a host of other problems, including family violence, despair, and alienation. Bringing participants together

Evaluating Empowerment Evaluations

Robin Lin Miller and Rebecca Campbell (2006) systematically examined 47 case examples of evaluations labeled "empowerment evaluation" published from 1994 through June 2005. They found wide variation among practitioners in adherence to empowerment evaluation principles and weak emphasis on the attainment of empowered outcomes for program beneficiaries.

> The larger picture that emerges from these data suggests that although many evaluation projects get labeled (and relabeled) as empowerment evaluations, frequently, these evaluations do not embody the core principles that are supposed to undergird empowerment evaluation practice . . . . Although empowerment evaluation advocates for the inclusion of program consumers in the evaluation, and it is they who ultimately are to be empowered, program recipients were seldom part of the empowerment evaluations, relative to what one might expect . . . Interpretations of empowerment evaluation in practice seem more narrowly focused on benefiting those who run and deliver programs. The goal of empowering citizens who are the beneficiaries of social programs has become less salient in cases of empowerment evaluation practice than has increasing the self-determining status of program staff members and managers and holding the program staff members accountable to funding institutions (Miller and Campbell 2006:314).

to establish mutual goals of support and identifying ways of evaluating (reality testing) goal attainment is a process of community development. The very process of working together on an evaluation has an impact on the group's collective identity and skills in collaborating and supporting each other. Participants also learn to use expert resources, in this case, the facilitating evaluator, but inquiry is democratized (IQREC 1997). One poverty program director explained to me the impact of such a process as she observed it:

> It's hard to explain how important it is to get people connected. It doesn't sound like a lot to busy middle class people who feel their problem is too many connections to too many things. But it's really critical for the people we work with. They're isolated. They don't know how the system works. They're discouraged. They're intimidated by the system's jargon. They don't know where to begin. It's just so critical that they get connected, take action, and start to feel effective. I don't know how else to say it. I wish I could communicate what a difference it makes for a group of poor people who haven't had many years of formal education to share the responsibility to evaluate their own program experiences, learn the language of evaluation, deal with data, and report results. It's very empowering.

Empowerment and Social Justice

The phrase "empowerment evaluation" can bridle. It comes across to some like a trendy, buzz word. Others experience it as oxymoronic or disingenuous. Still others find the phrase offensive and condescending. Few people, in my experience, react neutrally. Like the strategic planning term *proactive,* the word *empowerment* can create hostile reactions and may fall on hard times.

Empowerment carries an activist, social change connotation, as does a *related idea, using evaluation for social justice.* Vera, the main character in Nadine Gordimer's (1994) novel, *None to Accompany Me*, exclaims after a lengthy exchange about empowerment of South African blacks: "Empowerment, what is this new thing? What happened to what we used to call justice?" (p. 285). Perhaps Vera would have been pleased by the theme chosen by President Karen Kirkhart for the American Evaluation Association national conference in 1994 (the year after Empowerment Evaluation was the theme): "Evaluation and Social Justice."

The first prominent evaluation theorist to advocate valuing based on principles of social justice was Ernest House (1990b, 1980). He has consistently voiced concern for democratizing decision making (House 2004; House and Howe 2000). In that context, he has analyzed the ways in which evaluation inevitably becomes a political tool in that it affects "who gets what" (distributive justice). Evaluation can enhance fair and just distribution of benefits and responsibilities, or it can distort such distributions and contribute to inequality. In rendering judgments on programs, the social justice evaluator is guided by principles such as equality, fairness, and concern for the common welfare (Sirotnik 1990).

"Deliberative democratic evaluation" is an approach that attends to social justice by engaging the full range of stakeholder perspectives (the principle of inclusion) in dialogue and deliberation (House and Howe 2000, 1999; Ryan and DeStefano 2000). House has emphasized that such an inclusive and deliberative process reduces bias in findings:

> To be unbiased, evaluators might obtain stakeholder opinions from all relevant parties and process these values, views, and interests in systematic ways, thus balancing bias against bias in arriving at (relatively)

> unbiased conclusions. The process is analogous to enlisting different perspectives on panels to secure a balanced outcome or collecting diverse facts to obtain a correct perspective. (House 2004:222)

While House has emphasized the way in which deliberative democratic evaluation produces balanced results and reduces bias in findings, those stakeholders who participate in such deliberative and dialogic processes are learning to think evaluatively (Patton 2002b; Torres et al. 2000).

Feminist evaluation is another approach that emphasizes participatory, empowering, and social justice agendas (Podems 2005; Bamberger and Podems 2002; Seigart and Brisolara 2002). Feminist, social justice, and empowerment evaluations change the role of the evaluator from the traditional judge of merit or worth to a social change agent—or advocate (Greene 1997). Many evaluators surveyed by Cousins et al. (1996) were hostile to or at least ambivalent about whether participatory evaluation can or should help bring about social justice. Certainly, evaluators undertaking such an approach need to be comfortable with and committed to it, and such an activist agenda must be explicitly recognized by, negotiated with, and formally approved by primary intended users.

From a utilization-focused perspective, the important point about using evaluation processes to support social change is this: *Using evaluation to mobilize for social action, empower participants, and support social justice are options on the menu of evaluation process uses*. Since how these options are labeled will affect how they are viewed, when discussing these possibilities with primary intended users, evaluation facilitators will need to be sensitive to the language preferences of those involved.

Now, we turn to a conceptually different use of evaluation processes, evaluation for program and organizational development.

Evaluation for Program and Organization Development: Developmental Evaluation

The profession of program evaluation has developed parallel to the professions of management consulting and OD (organizational development) consultants advise on and facilitate a variety of change processes (Rothwell and Sullivan 2005; Carnevale 2002; Patton 1999a), including solving communications problems (Miller 2005); conflict resolution (Miall, Ramsbotham, and Woodhouse 2005; Angelica 1999); strategic planning (Bryson 2004a); leadership development (Crosby and Bryson 2005; Terry 2001; Schein 1985); organizational learning (Senge 2006; Hamilton et al. 2006; Preskill and Torres 1999); teamwork (Parker, Zielinski, and McAdams 2000); human resources (Dessler 2004); diversity training (Wildermuth 2005); shaping organizational culture (Schein 1989); and defining mission, to name but a few OD arenas of practice. Sometimes their methods include organizational surveys and field observations, and they may facilitate *action research* as a basis for problem solving (Whyte 1991; Schön 1987; Argyris, Putnam, and Smith 1985; Wadsworth 1984) or even evaluation (Patton 2002a; King 1995; Wadsworth 1993).

Program evaluation can be viewed as one approach on the extensive menu of organization and program development approaches. *Capacity building in evaluation can be a core part of more general organizational development initiatives and processes* (ECDG 2006; Kuzmin 2005; Horton and Mackay 2003; Horton et al. 2003). Evaluation's niche is defined by its emphasis on reality testing based on systematic data collection for improvement, rigorous assessment of outcomes, judging merit and worth, and generating knowledge about effectiveness. The processes of

evaluation support change in organizations by getting people engaged in reality testing, that is, helping them think empirically, with emphasis on specificity and clarity, and teaching them the methods and utility of data-based decision making.

For example, *evaluability assessment* (Wholey 1994; Smith 1989) has emerged as a process for evaluators to work with program managers to help them get ready for evaluation. It involves clarifying goals, finding out various stakeholders' views of important issues, and specifying the model or intervention to be assessed. From my perspective, this is really a fancy term that gives evaluators a credible niche for doing program and organizational development. Time and time again, evaluators are asked to undertake an evaluation only to find that goals are muddled, key stakeholders have vastly different expectations of the program, and the model that the program supposedly represents, that is, its intervention, is vague at best. In other words, the program has been poorly designed, conceptualized, or developed. To do an evaluation, the evaluator has to make up for these deficiencies. Thus, by default, the evaluator becomes a program or organizational developer. Rog (1985) studied the use of evaluability assessments and found that many of them precipitated substantial program change but did not lead to a formal evaluation. The programs realized, through the process of evaluability assessment, that they had a lot more development to do before they could or should undertake a formal evaluation, especially a summative evaluation. In such cases, the processes and logic of evaluation have impact on program staff quite beyond the use of findings from the assessment.

Mission-oriented evaluation is an organizational development approach that involves assessing the extent to which the various units and activities of the organization are consistent with its mission. For example, I evaluated the extent to which 550 grants made by the Northwest Area Foundation over 5 years were congruent with its mission. The board used that assessment at a retreat to review and then revise the organization's mission. The process of clarifying the foundation's mission with staff and board directors had at least as much impact as the findings (Hall 1992). I noted earlier in this chapter the very process of formulating a mission and goals so they can be evaluated will usually have an impact long before data are actually collected to measure effectiveness. In this way, evaluative questioning and facilitation can be understood as an organizational development activity.

Mission-oriented evaluation changes the unit of evaluation (unit of analysis) from the program to the organization. Many program and project evaluators have been self-limiting and missed organizational development opportunities by defining the primary unit of analysis for evaluation as the program or project. The question of organizational effectiveness can be quite different from program or project effectiveness. Let's look at some of the implications of changing the unit of analysis.

The Organization as the Unit of Analysis

First, the evaluation situation typically becomes more complex at the organizational level. There may be more stakeholders to deal with, more levels of stakeholders, and therefore greater challenges in sorting through various interests, interpersonal dynamics, and organizational politics. The environment surrounding and influencing an organization may also be more complex and dynamic compared with the environment of a single program. On the other hand, operating at the

organization level may increase the possibility of having impact by being able to deal directly with those who have the power to make changes.

Second, because programs and projects are usually embedded in larger organizational contexts, improving programs and projects may be linked to and even dependent on changing the organizations of which they are a part. For example, when evaluating the effectiveness of government programs, evaluators may need to examine, understand, and assess the ways in which being part of larger bureaucracies affect program and project effectiveness. Factors that can affect effectiveness include staff motivation, efficiency of program processes, and incentives to achieve outcomes, all of which are more determined at the organizational level than at the program or project level. Thus, improving programs may mean developing greater organizational effectiveness. This point deserves elaboration through an example and review of supporting data.

Systems Interconnections Between Levels and Units of Analysis

I once directed a knowledge-generating evaluation synthesis in search of overarching themes, patterns, and lessons learned that cut across the diverse experiences and evaluations of 34 separate Aid to Families in Poverty (FIP) programs supported by The McKnight Foundation in Minneapolis. The synthesis team consisted of five analysts. I had overall thematic integration responsibility while my four evaluation colleagues focused on integrating findings from the four cluster areas of the FIP initiative: employment, effective parenting/family stability, child care, and comprehensive poverty programs.

Our analytical process began with review of specific evaluation findings from separate projects. Then, we began looking for patterns across projects. We examined patterns of participant outcomes and project implementation. We found that effective projects shared some common characteristics, including the following:

- Effective staff are highly *responsive* to individual participants' situations, needs, capabilities, interests, and family context. In being responsive and respectful, they work to raise hopes and empower participants by helping them make concrete, intentional changes.
- Effective projects support staff responsiveness by being *flexible* and giving staff discretion to take whatever actions assist participants to climb out of poverty.
- Flexible, responsive projects affect the larger systems of which they are a part by *pushing against boundaries,* arrangements, rules, procedures, and attitudes that hinder their capability to work flexibly and responsively—and therefore effectively—with participants.

What is of interest for our purposes here is the extent to which we found an interconnectedness of these patterns at project, program, and organizational levels. An important breakthrough in our synthesis came when we understood that project and agency-wide organizational cultures were systemically interrelated. Systems consist of interdependent parts such that a change in one part affects other parts and the entire system. In examining FIP patterns across projects and host agencies, we found *that how people are treated affects how they treat others*:

> *responsiveness reinforces responsiveness, flexibility supports individualization, and empowerment breeds empowerment.*

Effective projects emphasized the importance of individualized, responsive, and respectful work with families in poverty. But we also found that staff generally could not (or would not) be individually responsive and supportive if they were part of organizational environments that were rigid and bureaucratic. We found that staff tended to treat participants the way they were treated as professionals within their organizations. If the program administration and environment were rigid, rules oriented, and punitive, the staff tended to be rigid, rules oriented, and blaming with participants. If the program environment and administration were flexible, responsive, nurturing, and supportive, the staff in that environment was more likely to interact with participants in ways that were responsive, nurturing, and supportive.

The systems connection between projects, programs, and organizations operated in both directions, however. Programs developed cultures that affected entire agencies and systems of which they were a part. The receipt of prestigious and relatively flexible philanthropic funding gave projects a certain degree of power, confidence, and autonomy such that they could make demands on their host organizations and cooperating units of government to be responsive and flexible. In many cases, when projects began implementation, they found that they could not do what needed to be done to assist participants because of agency or government rules and restrictions. Ordinarily, project leaders lacked the power or confidence to challenge system barriers to effectiveness. Many project leaders, however, with the open support of the independent philanthropic foundation funding the project, formulated strategies to overcome system barriers and thereby create a more flexible and responsive environment for their projects. Even as project staff were often called on to be advocates for families in poverty and help those families become assertive in overcoming system barriers they encountered, so also staff became assertive in working to change organizational and system barriers to program effectiveness. In some cases, then, projects changed the agencies in which they were housed, moving those host agencies toward greater flexibility and responsiveness. The evaluation synthesis, by highlighting these interconnections, led to further organizational development efforts. More directly related to the focus of this chapter, as our findings emerged in collaboration with leadership from these different organizations, their discussions, interactions, and reflections led many of them to change how they managed programs and projects. Well before we had produced formal findings and written a report, the processes of being engaged in the evaluation synthesis led those involved to begin thinking in different ways about how projects and programs related to their overall organizational cultures. In other words, being involved in the evaluation process, especially interpreting findings through interaction with leaders from other organizations, stimulated organizational development—a *process use* of the synthesis evaluation.

Overlapping Dimensions of Process Use: Organizational Development and Evaluative Thinking

The various kinds of process use discussed in this chapter are often found together. For example, infusing evaluative thinking into an organization's culture is one specific approach to overall organizational development and often involves participatory and collaborative evaluation as the

Organizational Development and Evaluative Thinking

The Five Most Important Questions You Will Ever Ask About Your Nonprofit Organization:

Peter Drucker died in 2005 at the age of 95. He invented the field of management consulting and after a highly successful career advising multinational corporations, he turned his attention to the challenges of increasing the effectiveness of not-for-profit organizations. While the bottom line for a business was monetary profit, he considered the bottom line for educational and nonprofit organizations to be changed lives. He founded the Drucker Foundation to promote increased effectiveness of not-for-profit organizations of all kinds. This led him to an organizational development framework that invited organizations to seriously engage with fundamentally evaluative questions. His classic work in this regard was a workbook entitled "*The Five Most Important Questions You Will Ever Ask About Your Nonprofit Organization."* Note the evaluative thinking that permeates his organizational development framework.

1. *What is our business (mission)?*
   - What are we trying to achieve?
   - What specific results are we seeking?
   - What are our major strengths?
   - What are our weaknesses?
2. *Who is our customer?*
   - Who are our primary customers?
   - Who are our supporting customers?
   - Have our customers changed?
   - Should we add or delete some customers?
3. *What does the costumer consider value?*
   - How well are we providing what our customers consider value?
   - How can we use what our customers consider value to become more effective?
   - What additional information do we need?
4. *What have been our results?*
   - How do we define results for our organization?
   - To what extent have we achieved these results?
   - How well are we using our resources?
5. *What is our plan?*
   - What have we learned and what do we recommend?
   - Where should we focus our efforts?
   - What, if anything, should we do differently?
   - What is my plan to achieve results for my group/responsibility area?
   - What is our plan to achieve results for the organization?

SOURCE: Peter Drucker (1993).

mechanism of development. Nowhere is this better illustrated than in the pacesetting organizational effectiveness initiative of the Bruner Foundation called the Evaluative Thinking in Organizations (ETHOS) project (Baker and Bruner 2006). Founded in 1963, the Bruner Foundation has made nonprofit evaluation methodologies and organizational effectiveness a major focus of its grant making. In 1996, the Bruner Foundation launched the Rochester Effectiveness Partnership (REP) in Rochester, New York, a pioneering project aimed at building evaluation capacity in local not-for-profit agencies through skill training, technical assistance, and participatory evaluation demonstration efforts. The initiative was founded on the premise that organizations that valued and used evaluation would become more effective, but that for organizations to learn to value and use evaluation, they would need guidance, training, technical assistance, and financial support. Between 1996 and 2003, 14 funding organizations and 32 social service provider organizations focused together on using evaluation to increase program and organizational effectiveness. The evaluations of this evaluation capacity-building effort showed substantial variation in the nature, quality, and utility of the program evaluations undertaken. Early on it became apparent that increasing the use of evaluation would require a supportive organizational culture and that building such a culture would necessitate organizational development and leadership support. Some organizations were open to this; others were not. It was clear that where program evaluation came to be valued, there were ripple effects within the larger organizations in which those programs resided. Thus, as the 7 years of the original capacity-building commitment was coming to an end, Beth Bruner and her colleagues turned their attention to this larger issue:

> While we had documented the spread or "ripple" of REP evaluation skills, we were curious to understand more about a possible relationship between increased evaluation capacity and the use of that capacity beyond the program level. In other words, could evaluation skills be applied to human resources, governance, communications and marketing as well as other organizational management areas? (Baker and Bruner 2006:xix)

This led, in 2004, to a new initiative, the ETHOS project, in which service provider partners in Rochester joined with funders and evaluators to inquire into the relationships between evaluation capacity, evaluative thinking, and organizational effectiveness.

> Evaluative Thinking was defined as a type of reflective practice that incorporated use of systematically collected data to inform organizational decisions and other actions. . . . The partners also clarified that evaluative thinking could be applied to many organizational functions (e.g., mission development/revision, human resources decision making) in addition to program development and service delivery. (Baker and Bruner 2006:34)

To support their inquiry, the evaluation and organizational partners reviewed various organizational assessment tools to develop an instrument that could be used to measure evaluative thinking practices within key organizational areas. The "ETHOS Assessment Tool" included multiple indicators in 15 organizational capacity areas, including Mission, Strategic Planning, Executive Leadership, Management Leadership, Governance,

Inquiry Question for Evaluative Thinking in Organizations (ETHOS) Project

1. What does evaluative thinking look like within various organizational capacity areas?
2. How are evaluative thinking and organizational effectiveness related?
3. What is needed to enhance and broaden evaluative thinking?

SOURCE: Baker and Bruner (2006:32).

Fund Development/Fund Raising, Evaluation, Program Development, Client Relationships, Communication and Marketing, Technology Acquisition and Training, Staff Development, Human Resources, Business Venture Development, and Alliances and Collaborations. The assessment tool was completed through site visits in which key decision makers from each organization were interviewed, for example, the executive director, representatives from upper- and mid-level management, line staff, representatives from the board of directors, and others as appropriate.

> Review of cross-site evaluative thinking findings showed that the partners' combined scores for each capacity area were relatively high. Respondents confirmed the presence of many indicators of evaluative thinking in their regular practices. There were some differences, however, in the scores. For example, the lowest combined score was 78 for Management Leadership and the highest combined score was 94 for Strategic Planning. This tells us that respondents saw more evidence of evaluative thinking in the way their organizations planned their actions than in the way they administered them. (Baker and Bruner 2006:40)

What is especially important for our purposes is that this Bruner Foundation initiative, for the first time, operationalized "evaluative thinking" and measured its manifestation in a variety of organizational functions across several organizations. Moreover, by using a participatory partnership research strategy, their inquiry into evaluative thinking offered evidence of cumulative process use, for the organizations that collaborated in the inquiry could point to ways that evaluative thinking had become more deeply infused in their organization's culture, could document specific changes they had made based on evaluation findings, could identify areas of enhanced organizational development, and reported increased understanding of and commitment to evaluative thinking for organizational effectiveness (Baker and Bruner 2006, especially pp. 49–57).

The combination of the Rochester Effectiveness Project followed by the ETHOS initiative took place over a decade with substantial financial support, training, technical assistance, and ongoing reinforcement. This contrasts dramatically with most such organizational learning and change initiatives that have a faddish, flavor-of-the-month feel, and lack sustainability and long-term commitment. Most organizational change initiatives fail precisely because they lack long-term commitment and follow through and prove unsustainable (Senge 1999). The successful case examples of infusing evaluative thinking into organizational culture we have presented in this chapter, IDRC and ETHOS, were both long-term, full-organization, leadership-engaged, and well-supported initiatives.

Intentional and Planned Process Use as an Option in Utilization-Focused Evaluations

Menu 5.1, presented earlier, summarizes the six primary uses of evaluation logic and processes discussed in this chapter. As

I noted in opening this chapter, any evaluation can, and often does, have these kinds of effects unintentionally or as an offshoot of using findings. *What's different about utilization-focused evaluation is that the possibility and desirability of using and learning from evaluation processes, as well as from findings, can be made intentional and purposeful—an option for intended users to consider building in from the beginning.* In other words, instead of treating process use as an informal ripple effect, explicit and up-front attention to the potential impacts of using evaluation logic and processes can increase those impacts and make them a planned purpose for undertaking the evaluation. In this way, the evaluation's overall utility is increased.

The six kinds of process use identified and discussed here—(1) infusing evaluative thinking into organizational culture, (2) enhancing shared understandings, (3) reinforcing interventions, (4) instrumentation effects and reactivity, (5) supporting participant engagement, and (6) developing programs and organizations—have this in common: They all go beyond the traditional focus on findings and reports as the primary vehicles for evaluation impact. As such, these new directions have provoked controversy.

Concerns, Caveats, and Controversies: Objections to and Abuses of Process Use

Just as evaluation findings can be misused, so too evaluation processes can be misused. Too much time can be spent worrying about evaluation to the detriment of direct service delivery. Evaluation can support interventions, but can also interfere with programs, as when so many forms have to be filled out or so much testing is done that time for direct work with those in need is significantly reduced. Resources and staff time devoted to evaluation are resources and staff time not available for working with clients (as staff, in frustration, will often readily point out). Indeed, evaluation can become its own source of goal displacement in which an organization becomes so enamored with its evaluation system that it's devoting more attention to that than to meeting client needs. One major foundation invited me in to look at its sophisticated process for capturing lessons. The foundation was, indeed, seriously engaged in analyzing and reflecting on what was working and not working; evaluative thinking clearly permeated the foundation's documents, plans, staff meetings, and governance sessions. The problem was that it was all thinking and no doing. They were congratulating themselves for openly acknowledging failures, but they have been documenting the same failures over and over again for several years and nothing seemed to have significantly changed. They were heavy into process use—*but had neglected findings use*. They were thinking and talking a lot about evaluation, but, in fact, they weren't adept at actually using findings. All the attention to and rhetoric about evaluation, and their precious status as a knowledge-generating learning organization, had disguised the fact that they were quite ineffective in actually using evaluation findings. *Process use is no substitute for findings use. Process use should enhance findings use.*

In addition to the classic concern that too much attention to process may undercut attention to outcomes (in this case, findings use), six other objections—closely interrelated, but conceptually distinct—arise most consistently when process use is proposed as an evaluation option.

1. *Definitional objection.* Evaluation should be narrowly and consistently defined

in accordance with the "common sense meaning of evaluation," namely, "the systematic investigation of the merit or worth of an object" (Stufflebeam 1994:323). Anything other than that isn't evaluation. Adding terms such as empowerment or collaborative to evaluation changes focus and undermines the essential nature of evaluation as rendering judgment.

2. *Goals confusion objection.* The goal of evaluation is to render judgment. "While . . . 'helping people help themselves' is a worthy goal, it is not the fundamental goal of evaluation" (Stufflebeam 1994:323). Goals such as organizational development, evaluation capacity building, and infusing evaluative thinking into the organization's culture are worthy undertakings, but are not evaluation.

3. *Role confusion objection.* Evaluators as people may play various roles beyond being an evaluator, such as training clients or helping staff develop a program, but in taking on such roles, one moves beyond being an evaluator and should call the role what it is, for example, trainer or developer, not evaluator.

> While one might appropriately assist clients in these ways, such services are not evaluation. . . . The evaluator must not confuse or substitute helping and advocacy roles with rendering of assessments of the merit and/or worth of objects that he/she has agreed to evaluate. (Stufflebeam 1994:324)

Scriven (1991a) has been emphatic in arguing that being able to identify that something is or is not working (an evaluator's role) is quite different from knowing how to fix it or improve it (a designer's role).

4. *Threat to data validity objection.* Quantitative measurement specialists teach that data collection, for the results to be valid, reliable, and credible, should be separate from the program being evaluated. Integrating data collection in such a way that it becomes part of the intervention contaminates both the data and the program. Designing instruments to contribute to learning and reflection undercut their validity to measure program processes and outcomes accurately.

5. *Loss of independence objection.* Approaches that depend on close relationships between evaluators and other stakeholders undermine the evaluator's neutrality and independence. "It's quite common for younger evaluators to 'go native,' that is, psychologically join the staff of the program they are supposed to be evaluating and become advocates instead of evaluators" (Scriven 1991a:41). This can lead to overly favorable findings and an inability to give honest, negative feedback.

6. *Corruption and misuse objection.* Evaluators who identify with and support program goals, and develop close relationships with staff and/or participants, can be inadvertently co-opted into serving public relations functions or succumb to pressure to distort or manipulate data, hide negative findings, and exaggerate positive results. Even if they manage to avoid corruption, they may be suspected of it, thus undermining the credibility of the entire profession. Or these approaches may actually serve intentional misuse and foster corruption, as Stufflebeam (1994) worries,

> What worries me most about . . . empowerment evaluation is that it could be used as a cloak of legitimacy to cover up highly corrupt or incompetent evaluation activity. Anyone who has been in the evaluation

> business for very long knows that many potential clients are willing to pay much money for a "good, empowering evaluation," one that conveys the particular message, positive or negative, that the client/interest group hopes to present, irrespective of the data, or one that promotes constructive, ongoing, and nonthreatening group process. . . . Many administrators caught in political conflicts would likely pay handsomely for such friendly, nonthreatening, empowering evaluation service. Unfortunately, there are many persons who call themselves evaluators who would be glad to sell such service. (P. 325)

Exhibit 5.5 summarizes contributions and concerns about different kinds of process use. The concerns have sparked vigorous debate (e.g., Fetterman 1995). In Chapter 14 on the politics and ethics of utilization-focused evaluation, I'll address these concerns with the seriousness they deserve. For the purpose of concluding this chapter, it is sufficient to note that the utilization-focused evaluator who presents to intended users options that go beyond narrow and traditional uses of findings has an obligation to disclose and discuss objections to such approaches. As evaluators explore new and innovative options, they must be clear that *dishonesty, corruption, data distortion, and selling out are not on the menu*. Where primary intended users want and need an independent, summative evaluation, that is what they should get. Where they want the evaluator to act independently in bringing forward improvement-oriented findings for formative evaluation, that is what they should get. But those are no longer the only options on the menu of evaluation uses. New participatory, collaborative, intervention-oriented, and developmental approaches are already being used. The utilization-focused issue is not whether such approaches should exist. They already do. The issues are understanding when such approaches are appropriate and helping intended users make informed decisions about their appropriateness.

Crossing Borders, Crossing Boundaries

The theme of the 2006 International Evaluation Conference in Toronto was "Crossing Borders, Crossing Boundaries." In the spirit of that theme, I close this chapter by updating the Sufi story with which this chapter began. Nasrudin had fooled customs officials by smuggling donkeys, which were in plain sight, when they were forever searching him for presumed contraband not in plain sight. Having retired wealthy from smuggling donkeys, he encountered a former customs officials to whom he confessed what he had been smuggling. The man then asked, "And what are you doing these days?"

"Smuggling ideas and new ways of thinking across closed borders," Nasrudin replied. "I get paid more for a good idea than I ever got paid for a donkey. Though good ideas are harder to come by, they're easier to carry across borders." (See illustration on the next page.)

The idea that evaluation can be useful as a way of thinking is not always a welcome or credible idea. Sometimes it has to be introduced to closed minds and rigid compliance-based organizations with cunning and stealth. And sometimes not. Knowing the difference is the challenge of situational responsiveness, the topic of Chapter 6.

Follow-Up Exercises

1. Identify a real program. Describe the program's activities, intended outcomes, and its context. Choose one kind of *Process Use* from Menu 5.1 and describe how that process could be applied to an evaluation of this program. What would be the benefits of such evaluation process use? What disadvantages or difficulties can be anticipated?

2. Locate someone who has experienced a powerful learning experience or a major change in their life and is willing to be interviewed. Interview that person about the major change experienced and how that change has affected the person's current life—actions, behaviors, attitudes, feelings, and priorities. Probe about the nature of the change or learning experience, the situation that led to the change or learning, and the factors that affected what happened. At the end of the interview, ask questions about the experience of being interviewed. Did the person come to any new understandings about the learning experience or change that was the focus of the interview? Probe for reactions to being interviewed and reflecting on the questions asked. Analyze those reactions as instrumentation effects and discuss this kind of reactivity as an example of process use.

3. Conduct an Internet search of the term *participatory evaluation*. Examine the variations in what is meant by and described as "participatory evaluation." Discuss the implications of these variations for evaluation practice. How does utilization-focused evaluation deal with such variations in definition, meaning, and practice?

4. Write out a script for explaining *process use* and process use options to a philanthropic foundation funder who might be interested in funding evaluations. In your own words, with a specific person in mind, tell them about process use.

5. Find your own analogy for describing and explaining process use. This chapter opened and closed with a Sufi tale that illustrated aspects of process use. I also offered hiking the Grand Canyon as an analogy for process use (see page 155). Come up with your own analogy for process use.

EXHIBIT 5.5

Process Use Contributions and Concerns

Type of Process Use	*Contributions*	*Concerns, Challenges, and Caveats*
1. Infusing *evaluative thinking* into the organizational culture	Evaluation becomes part of the organization's way of doing business, contributing to all aspects of organizational effectiveness. People speak the same language, share meanings, and priorities. Reduces resistance to evaluation.	Requires consistent leadership, ongoing training, and reinforcement (rewards). The rhetoric is easy, but actually internalizing evaluative thinking is difficult. Can increase conflict between those who "get it" and those who don't.
2. Enhancing shared understandings	Gets everyone on the same page; supports alignment of resources with program priorities.	Can force premature closure before alternatives are fully considered; those with the most power may impose their perspective on the less powerful.
3. Supporting and reinforcing the program intervention	Enhances outcomes and increases impact; increases the value of evaluation	Complicates attribution; the effects of the program become intertwined with the effects of the evaluation, in effect making the evaluation part of the intervention.
4. Instrumentation	What gets measured gets done. Focuses program resources on priorities. Measurement contributes to participant learning.	Measure the wrong things, the wrong things gets done. Goal displacement, where what can be measured becomes the program's goals. Corruption of indicators, especially where the stakes become high.
5. Increasing engagement, self-determination, and ownership	Makes evaluation especially meaningful and understandable to participants; empowering.	Can reduce the credibility of the evaluation to external stakeholders; comingles evaluation purposes with empowerment agendas, potentially undermining both.
6. Program and organizational development	Capacity building; long-term contribution beyond specific findings; enhances ongoing adaptability.	Evaluator serves multiple purposes and plays multiple roles, potentially confusing the evaluation role; evaluator develops a close relationship with program decision makers, raising questions about the evaluator's independence.