



Assess and Build Program and Organizational Readiness for Utilization- Focused Evaluation

Reality isn't the way you wish things to be, nor the way they appear to be, but the way they actually are.

Robert J. Ringer,
American entrepreneur and author

The phrase “reality testing” implies that being “in touch with reality” cannot simply be assumed. When individuals lose touch with reality, they become dysfunctional and, if the distortions of reality are severe, they may be referred for psychotherapy. Programs and organizations can also “lose touch with reality” in the sense that the people in those programs and organizations are operating on myths and behaving in ways that are dysfunctional to goal attainment and ineffective for accomplishing desired outcomes. As the introductory chapter emphasized, program evaluation provides a mechanism for finding out whether what’s hoped for is, in fact, taking place—a form of reality testing.

Getting people ready for evaluation means helping them commit to reality testing. Research on *readiness for evaluation* has shown that “valuing evaluation” is a necessary condition for evaluation use (Patton, 2008, pp. 42–46). Valuing evaluation cannot be taken

for granted. Nor does it happen naturally. Users' commitment to evaluation is typically fragile, often whimsical, and must be cultivated like a hybrid plant that has the potential for enormous yields, but only if properly cared for, nourished, and appropriately managed. When beginning a new program evaluation or entering an organization to improve an existing evaluation system, it is wise to begin with an assessment of existing attitudes toward and behaviors related to evaluation use. Here is some of what to watch for.

Naysayers. Some people seem to be naturally negative—about most everything. You probably know someone like this. Whenever an idea comes up that involves change, their instant reaction is: Won't work here.

The complacent. These people belong to the *things are okay as is* crowd. If it's not broken, why fix it? And from their perspective, nothing involving them is ever broke. They've constructed their own comfortable worlds built on untested assumptions and unexamined beliefs. Evaluation is a threat to such people. Reality testing will only upset things. "Why bother?" they ask.

The compliant. Many program administrators experience evaluation only as something mandated by funders so they've developed a compliance mentality. The compliant-oriented say: "Just tell us what we have to do and we'll do it." They don't want to think about it or engage with it, they just want to meet the mandate. They know that money comes with strings attached. They don't like it, but they accept it. Ironically, mandated evaluations can actually undercut utility. A career internal evaluator with 25 years experience once told me: "The most pervasive problem we deal with, particularly in our relations with state and federal agencies, is submitting mandated reports that no one cares about, no one reads, and we never get any feedback on. It is constantly discouraging and frustrating."

The cynics. These are the "been there, done that" folks. They have evaluation horror stories to share. Waste of time and money. Academic. Esoteric. Irrelevant. Useless. These are old-timers who take pride in having survived lots of what they consider bad evaluations—and lived to tell about it. And tell about it they do, to any and all who will listen.

The frightened and anxious. These tend to be people who confuse program evaluation with personnel evaluation and are worried that THEY are being evaluated. They also know some of the program's weaknesses but hope such blemishes will stay hidden from outside eyes. Now they fear exposure and imagine ruin.

The enthusiastic. Some people hunger for learning. They want to improve. They may even be data junkies—can't get enough of it. These can be allies, but ironically, some of the

most enthusiastic adherents of evaluation are adherents of the idea but not the practice. They believe that evaluation is important and like the idea of gathering data, but they haven't thought through the possibility that findings may show a need for change.

The sophisticates. This person values data, is open to learning and feedback, understands that data must be interpreted to take action, and is prepared to engage in the evaluation process from design to interpretation and use. Sophisticates can be strong allies. They can also create conflicts by insisting on their way of doing things and acting with disdain toward those stakeholders who are less sophisticated.

The offended. These are sincere, committed, hard-working and dedicated true believers who are doing everything they know how to do to help people. They work long hours under difficult conditions with limited resources and engage daily with the afflictions that beset humanity's least privileged and most neglected: people ravaged by extreme poverty; those suffering from HIV/AIDS; the hungry and homeless; abused children; and victims of war, natural disasters, oppression, and all the many varieties of humankind's inhumanity toward fellow humans. Then some evaluator comes along and asks: "How do you know that you're doing good?" They are offended by the question. They are even more offended by the fact that resources are being spent on evaluation that could be used instead to feed the hungry, house the homeless, or comfort the afflicted: "Who the hell do you think you are coming in here with your surveys and data forms? We're dealing with life and death here. Get out of my face." At such moments it is worth remembering that they are right. Funds spent on evaluation are funds not spent on direct help to those in need. Evaluation's contribution is indirect. The justification for our inquiry is the belief and hope that we can help make them more effective. But they must join us in that belief. And we must be able to make the case that we have something to offer that will justify the costs of evaluation.

The politicians. These are folks whose attention is riveted on which way the political winds are blowing. They view data and evaluation findings like they view everything else—ammunition for political combat. They are attuned to the political power of a program's constituencies. If powerful constituents want the program, or if more is to be gained politically by support for rather than opposition to the program, then the program is judged worthwhile; no other evidence of program effectiveness is needed, though data may be sought to support this predetermined political judgment. Political vested interests are one reason it is so difficult to terminate government-funded programs and agencies. Programs develop constituencies whose vested interests lie in program continuation. The driving interest of *politicos* is to direct program funds where they count politically, not where they will be used most effectively. *Politicos* judge programs as effective as long as

they serve powerful interests. Empirical evaluation findings are of interest only insofar as they can be manipulated for political and public relations purposes. Politicos emphasize findings that support their political interests and dismiss, ignore, or distort findings that run counter to their political commitments and vested interests. Evaluators have to be politically astute to deal with *politicos*.

Now, of course, these are all caricatures and stereotypes and don't do justice to the great diversity and individuality of those wonderful folks we call "stakeholders." The point is to pay attention to the characteristics of the stakeholders who will be involved with an evaluation. Program and organizational cultures consist of a mix of types of people. The blend of perspectives in any given program gives rise to an organizational culture that will affect how evaluation is undertaken. Taking time at the beginning to understand and assess the program and organizational culture is the beginning of the utilization-focused evaluation process. Methods-driven evaluators want to cut to the chase and get right to work designing the evaluation and generating instruments. Utilization-focused evaluators, in contrast, begin by figuring out the context within which the evaluation will unfold and learning about the people within that context.

Learning About and Assessing the Evaluation Context

Evaluations are funded and commissioned by someone or some group. Those who fund and commission the evaluation are the evaluator's client. However, they are not the only people who will be affected by the evaluation or use evaluation findings. A government agency or philanthropic foundation may mandate and fund an evaluation, but the people running the program may be expected to make primary use of the evaluation for program improvement. In other cases the evaluation may be done for accountability purposes or to make a major decision about future funding or expansion of the program. Program participants have a stake in how the evaluation is conducted and what is done with findings. In later chapters we'll discuss in depth and detail alternative evaluation purposes and uses. At this point, we want to begin by getting the clients' perspective on the evaluation and assessing the evaluation context.

The first step is typically reviewing important documents and meeting with the key clients. This is often followed by interviewing some other key stakeholders, either individually or sometimes in a focus group. This is a chance to meet key people (and for them to meet the evaluator), learn more about the program's history, find out about important political issues, get a sense of the organization's culture, and interact around why the evaluation is being conducted and how it is perceived. This is also a chance to assess how committed to testing reality the client and key stakeholders are.

Because evaluators have typically internalized the value of data-based reality testing, it is easy to assume that others share this perspective. But a commitment to examine beliefs and test actual goal attainment is neither natural nor widespread. Evaluators who ignore the threatening nature of reality testing and plow ahead with instrument design and data collection in the hope that knowledge will prevail are engaged in their own form of reality distortion. Utilization-focused evaluators, in contrast, work with intended evaluation users to help them understand the value of reality testing and buy into the process, thereby reducing the threat of evaluation and resistance (conscious or unconscious) to evaluation use. A common error made by novice evaluators is believing that because someone has requested an evaluation or some group has been assembled to design an evaluation, the commitment to reality testing and use is already there. Quite the contrary, these commitments must be engendered (or revitalized if once they were present) and then reinforced throughout the evaluation process. Utilization-focused evaluation makes this a priority.

Once you have met key people, started learning about the organizational and political context for the evaluation, and have a sense of the issues the evaluation will need to address, it is time to move to the next level of formal stakeholder engagement. The remainder of this chapter presents some exercises that can be used to assess readiness for evaluation and give an evaluator a feel for the people who are being asked to undertake an evaluation and the program culture within which it will take place. Strategies for engagement flow from what is learned in these opening exercises that take place at an evaluation launch workshop.

Formally Launching the Evaluation: Exercises to Help Assess and Build Readiness for Evaluation—and Get the Process Started

To begin an evaluation, I like to assemble key people for a half-day workshop (sometimes a full day) aimed at laying the groundwork (tilling the program and organizational soil) for reality testing. People invited to this opening workshop include key program and organizational leadership, staff, funders, board members, and long-time program participants or graduates who are knowledgeable about the program. I'll say a lot more about selecting and involving these various stakeholders later. Right now I want to share examples of 10 exercises that help get people ready for reality testing, that is, begin the process of creating an evaluative mindset among those who will be involved in, affected by, and use the evaluation. Menu 1.1 summarizes these 10 exercises and the situations for which each is especially appropriate. The remainder of this chapter elaborates on the rationale for and what is involved in using each exercise.

MENU 1.1

Menu of Exercises to Assess and Facilitate Stakeholder and Program Readiness for Utilization-Focused Evaluation

<i>Exercises to assess & build readiness for utilization-focused evaluation</i>	<i>Situation for which this exercise is particularly appropriate</i>
1. <i>Baseline assessment of evaluation use:</i> How are data currently collected being used? How have past evaluations been used?	Organizations already engaged in some evaluation and/or ongoing data collection.
2. <i>Baseline associations with and perceptions of evaluation:</i> “What comes to mind when you see the word EVALUATE?”	Use with a group of stakeholders brought together to launch a new evaluation effort; surfaces the “baggage” that people bring to the new initiative from past experiences.
3. <i>Create a positive vision for evaluation:</i> If evaluation was really useful and actually used here, what would that look like?	Use to help a group move from evaluation anxiety and resistance to focus on the potential benefits of evaluation. Creates readiness for evaluation by focusing on use and creating a shared group understanding of and commitment to use the evaluation.
4. <i>Assess incentives for and barriers to reality testing and evaluation use in their own program culture.</i>	Once a group has a general sense of evaluation’s potential utility, this exercise takes the next step of getting concrete about what will need to occur with this program context to make evaluation useful.
5. <i>Engender commitment to reality-testing:</i> Are you willing to take a close look at whether what you think is happening in this program is actually happening, and whether what you hope it is accomplishing is actually being accomplished?	These questions are useful at the moment of commitment after some basic groundwork has been laid. These questions implicitly ask those involved: <i>Are you ready to get serious?</i>
6. <i>Grounding any specific evaluation in the professional standards of evaluation: utility, feasibility, propriety, and accuracy.</i> Give a copy of the	This exercise is useful for groups with little knowledge about the profession of evaluation. It helps evaluators position themselves and their work within a larger

<i>Exercises to assess & build readiness for utilization-focused evaluation</i>	<i>Situation for which this exercise is particularly appropriate</i>
standards to stakeholders involved with the evaluation and give them a chance to discuss what they see in the standards.	context of professional standards of excellence.
7. <i>Establishing project-specific evaluation norms and standards:</i> In addition to the evaluation profession's standards (# 6), what other values will be important in conducting this evaluation?	This exercise is useful with a group of particularly diverse stakeholders who come from different backgrounds and work in diverse professions with their own standards. This exercise offers the group a chance to articulate their own context-specific standards.
8. <i>Sharing good ideas that haven't worked out in practice.</i> This can be done in small groups where participants share their examples and how they came to see that what they thought was a good idea at the time didn't work out so well.	This exercise illustrates evaluative thinking in practice. It gives participants a chance to learn more about each other and start to create trust by sharing things about themselves and hearing each others' stories.
9. <i>Generate evaluation metaphors and analogies.</i> Have participants share some activities that they engage in (like fishing, sports, artistic creations, exercise, cooking, gardening) and look at the evaluation issues that are part of such activities. Or have them create evaluation metaphors around objects (chalkboards, light-bulbs, coffee cups).	This exercise engages participants creatively with each other while having some fun by sharing the associations they conjure up related to evaluation. This doesn't take much time to do and carries the message that everyone brings some evaluation expertise and experience to the process.
10. <i>Generating questions for evaluation:</i> What don't you know that if you did know would make a difference to your engagement with this program? What do you want to find out through this evaluation?	This is especially useful with groups that are expecting the evaluator to determine the evaluation questions. Engaging a group in this exercise sends the message that <i>their</i> questions and issues will focus the evaluation. This moves them from general evaluation principles to getting specific about the evaluation that is to be designed here and now.

1. Baseline assessment of evaluation use

Baseline data are critical in any evaluation. You can't tell how much participants have benefitted from a program if you don't know what their status was at the beginning. Administering a pretest before a program begins is a common way to establish baseline. Intake forms gather information when participants enter a program and are a source of baseline data. Needs assessments used to design a program typically include baseline indicators. Comparing program outcomes against an initial baseline is fundamental evaluation analysis.

The same logic applies to increasing evaluation use. What is the baseline of evaluation use in the program? Even if the program itself is new, it is likely that the people running and staffing the program have prior evaluation experiences and perspectives based on those experiences. Thus, a good place to begin a program and organizational assessment is to learn the organization's past and current evaluation and management information system practices. This can be done in interviews with key people who know the program, in focus groups, or in small-group discussions at the evaluation launch workshop. Questions to explore include: What information is currently collected and how is it used? What are examples of evaluations that you consider useful? What made them useful? What are examples of evaluations that were not used? What were the barriers to use? How would you say evaluation is viewed here? What are key factors to take into account in designing an evaluation to enhance its utility?

2. Baseline associations with and perceptions of evaluation

As a simple opening exercise at an evaluation launch workshop, I like to begin by asking participants to share words and feelings they associate with evaluation, then we explore how their perceptions they've brought with them may affect their expectations about the evaluation's likely utility. I write the word "EVALUATE" on a flip chart and ask those present to free-associate with the word: "What comes to mind when you see the word EVALUATE?" They typically begin slowly with synonyms or closely related terms: *assess, measure, judge, rate, compare*. Soon someone calls out "*waste of time*." Another voice from the back of the room yells: "*crap*." The energy picks up and more associations follow in rapid succession: *budget cuts, downsize, politics, demeaning, pain, fear*. And inevitably, the unkindest cut of all: "USELESS."

Clearly, evaluation can evoke strong emotions, negative associations, and genuine fear. People carry with them into new experiences the emotional baggage of past experiences. To ignore such perceptions, past experiences, and feelings people bring to an evaluation is like ignoring a smoldering dynamite fuse in the hope it will burn itself out. More likely, unless someone intervenes and extinguishes the fuse, it will burn faster and eventually explode. Many an evaluation has blown up in the face of well-intentioned evaluators because they rushed into technical details and methods decisions without establishing a solid foundation

for the evaluation in clear purposes and shared understandings. To begin, both evaluators and those with whom we work need to develop a shared definition of evaluation and mutual understanding about what the process will involve, and in so doing, acknowledge anxiety and fears.

Whether evaluations are mandated or voluntary, those potentially affected by the evaluation may approach the very idea with trepidation, manifesting what has come to be recognized by experienced evaluations as “evaluation anxiety”—or what I jokingly refer to with clients as a clinical diagnosis of *pre-evaluation stress syndrome*. But the fear is often serious and needs to be acknowledged and managed. Signs of extreme evaluation anxiety include “people who are very upset by, and sometimes rendered virtually dysfunctional by, any prospect of evaluation, or who attack evaluation without regards to how well conceived it might be” (Donaldson, Gooler, & Scriven, 2002, p. 262).

Moreover, there are genuine reasons for people to fear evaluation. Evaluations are sometimes used as the rationale to cut staff, services, or entire programs. Poorly done evaluations may misrepresent what a program has done and achieved. Even when an evaluator has done a good job, what gets reported in the news may be only the negative findings—or only the positive findings—rather than the balanced picture of positives and negatives, strengths and weaknesses, that was in the full evaluation report. These things happen. There’s no point in denying them. Evaluations can be well done or poorly done, useful or useless. By acknowledging these realities, we can begin the discussion of how, for this evaluation in this time and place, what do we have to do to undertake an evaluation that will be useful, credible, meaningful, and fair?

3. Creating a positive vision for evaluation

Because of bad past experiences and understandable fears about being unfairly judged, people don’t necessarily or naturally think about what evaluation can offer if it is well done. So, having established a baseline of past and current evaluation practices in the program and organization, and having gotten out in the open the perceptions and concerns of those who will be involved in and affected by the evaluation, we can begin the process of creating a positive vision for evaluation. In small groups, I ask people to create scenarios in response to the questions:

If evaluation was really useful and actually used here, what would that look like?

What could take place that would tell you that the evaluation was successful?

What would evaluation success look like?

The notion here, which is true in fields as disparate as sports, politics, business, and the arts, is that people are more likely to achieve success if they can envision success in a concrete

way. That's also true, by the way, for participants and staff in programs, which is another reason evaluators push for clarity about intended program outcomes. We're applying the same logic here to the evaluation: What are the positive intended outcomes for the evaluation? This exercise aims at transforming a skeptical, hostile, or dubious program culture into one that is at least open to the possibility of engaging in evaluation being a positive and useful experience.

4. Assess incentives for and barriers to reality testing and evaluation use in their own program culture

Having created together a positive vision of evaluation, the next step is to have the group identify what needs to happen to realize that vision as well as identify any barriers that will need to be overcome. As I work with intended users to agree on what we mean by evaluation and engender a commitment to use, I invite them to assess incentives for and barriers to reality testing and information use in their own program culture. Barriers typically include fear of being judged, cynicism about whether anything can really change, skepticism about the worth of evaluation, concern about the time and money costs of evaluation, and frustration from previous bad evaluation experiences, especially lack of use. As we work through these and related issues to "get ready for evaluation," the foundation for use is being built in conjunction with a commitment to serious and genuine reality testing.

Sometimes, indeed often, there are issues that people are reluctant to discuss openly. There may be huge personality conflicts with bosses or funders. There may be suspicions that key decisions have already been made and the evaluation is just a pretense to justify and put window-dressing around those already-made decisions. Some may know of irregularities, even illegalities, that the evaluation could expose. A key person who is supposed to lead the evaluation internally may be viewed as incompetent and has been assigned to lead the evaluation as punishment or because the administrator just needs to give that person something to do. How do you find out about these beneath-the-surface issues that may derail an evaluation? Here's what I do.

I give everyone the following piece of paper:

CONFIDENTIAL FEEDBACK

On this sheet of paper, please tell me anything you think I should know about what goes on in this program or organization that could affect how the evaluation is conducted and used. As an outsider, what do you want me to know that may not surface in open group discussions? This information is for **my eyes only**. Please return this to me personally. Don't put your name or any identifying information on this paper. However, if you don't want to write down your concerns but wish to talk to me confidentially, please provide your contact information.

I always alert the evaluation client and those organizing the launch workshop that I'll be doing this. Sometimes I learn little from asking for this feedback. At other times amazing, important, and disturbing things emerge. This is also a trust-building exercise and a chance to surface concerns early when strategies can still be developed to deal with issues that might derail the evaluation if they are not addressed. I typically invite this feedback just before a break or at the end of the session so that I can collect the feedback myself as people are leaving the room. Once collected, confidentiality must be protected absolutely. This is context information to help navigate the sometimes treacherous terrain of programs, organizations, and communities. As Spanish novelist Miguel de Cervantes had Don Quixote proclaim: "Forewarned, forearmed; to be prepared is half the victory."

5. Engender commitment to reality-testing

The preceding exercises have embedded within them getting a program group to begin thinking evaluatively. It gets them thinking about baselines and desired results and how to move from the baseline to desired results. The next step is to deepen the commitment to reality testing.

To do so, I like to appeal to what I presume to be a desire to be effective. I don't encounter many people who get up in the morning wanting to be ineffective. Indeed, the desire to be effective is so strong that it is often the source of evaluation anxiety. What if I've been wasting my time? What if what we've been doing isn't helping people? What then?

What then, indeed? The evaluation answer: Change the program to become more effective. So, engendering a commitment to evaluation use often involves engendering an openness to change. That openness can start with a commitment to examine reality. I ask: Are you willing to take a close look at whether what you think is happening in this program is actually happening, and whether what you hope it is accomplishing is actually being accomplished? Are you willing and ready to engage in serious reality testing?

Because evaluation use is so dependent on the commitment to reality testing, evaluators need ways to cultivate that commitment and enlarge the capacity of intended users to undertake the process. This means engaging program staff, managers, funders, and other intended users in examining how their beliefs about program effectiveness may be based on selective perception, predisposition, prejudice, rose-colored glasses, unconfirmed assertions, or simple misinformation. I often share evidence from psychology showing that we all tend to distort reality. For example, social scientist Yi-Fu Tuan (2000) has studied the human propensity toward escapism as a way of avoiding facing reality and even defines human beings as "animals congenitally indisposed to accept reality as it is" (p. 6).

One way we distort reality is through the "confirmation bias," the tendency to look for evidence to confirm our preconceptions and biases. If a program director believes his program is effective, he'll look for evidence of effectiveness and discount or ignore evidence of ineffectiveness. If a philanthropic foundation executive believes she is doing good, she'll find evidence of doing good wherever she looks. Our brains look for patterns, and once we

decide there is a pattern, we quite naturally and persistently look for confirming evidence. At the same time, we tend to dismiss contradictory evidence as unreliable or invalid. Systematic and balanced evaluation aims to counter our human tendency to distort reality.

Another barrier to reality testing is overconfidence in our judgments. Study after study shows that we tend to be overconfident about our ability to make informed and objective decisions. For example, the work of Kahneman and Tversky (2000b) on “prospect theory” has established that people have an aversion to loss that is about twice the desire for gain; in essence, we fear bad news and failure twice as much as we hope for good news and success. Brain research is helping us understand these proclivities, but culture and socialization also play key roles. The research of psychologists David Dunning and Justin Kruger (1999) brings irony to this picture of human reality distortion. They found that people who perform poorly are typically quite confident about their competence. In fact, they’re often more confident than those who actually perform well. The title of their research report sums up the problem quite succinctly: “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments.” Incompetent people also lack the skill to recognize competence in others.

A classroom experiment by behavioral economist Richard Thaler provides yet another glimpse of common reality distortion. At the beginning of his course on decision making, he asks students to anonymously write down where they’ll rank in the final grading. He reports that none of his 125 MBA students thought they would finish in the bottom half of the class. “Obviously, half of them were wrong,” he adds. “We all think we are pretty good at sizing things up, just the same way we all think we are good judges of character. We should know that we really are all hopeless” (quoted in Peltz, 1999, p. 105).

In addition to our tendency to deceive ourselves and distort reality, the irony of living in the information age is that we are surrounded by so much misinformation and act on so many untested assumptions. By putting intended users in touch with how little they really know, and how flimsy is the basis for much of what they think they know, we are laying the groundwork for evaluation use. We are, in fact, identifying that there are useful things to be found out, establishing that systematic inquiry processes are needed to counter our tendencies to distort reality, and creating the expectation that testing reality will be valuable, not just an academic or mandated exercise. In short, we are establishing psychological and organizational openness to and readiness for serious evaluation.

6. Grounding any specific evaluation in the professional standards of evaluation

As I work with key stakeholders toward a shared understanding of evaluation and a clear commitment to reality testing and evaluation use, I look for opportunities to review the development of program evaluation as a field of professional practice and present the standards for and principles of evaluation. (See the introductory chapter, page 5.). The *Standards* call for evaluations to be useful, practical, ethical, accurate, and accountable (Joint Committee on Standards for Educational Evaluation, 2010). Sharing the standards

communicates to primary intended users that evaluation has developed into an established profession—and that those engaged in evaluation have an obligation to act in accordance with professional standards and principles, including priority attention to utility.

I always give participants in the evaluation launch meeting or workshop a copy of the standards and ask them for their reactions. What stands out to you? What comes across to you? I explain to them that these are the professional standards for which I am accountable. We begin with my accountability before dealing with theirs. We then discuss the implications of the standards for the particular evaluation we are working on.

Few nonevaluators are aware of the field's professional associations, conferences, journals, standards, and principles. By associating a particular evaluative effort with the larger profession, you can elevate the status, seriousness, and meaningfulness of the process you are facilitating, and help the primary intended users understand the sources of wisdom you are drawing on and applying as you urge them to attend carefully to utilization issues from the start.

Let me make special mention of what may well be the most challenging of all the standards: the cost-effectiveness standard. In opening this chapter, I offered caricatures of various stakeholder types one may encounter. One of those was *the offended*, people working very hard on difficult problems with limited resources. They question the value of spending scarce funds on evaluation when the resources to meet pressing human needs are so few. One of the original standards addressed this concern:

Cost Effectiveness—The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified. (Joint Committee, 1994, F3)

The justification is that the findings will be used to make improvements in programs and inform important decisions that will ultimately help more people. No standard is more important—or more challenging to meet.

7. Establishing project-specific evaluation norms and standards

Introducing the evaluation profession's standards to stakeholders can offer an opportunity to generate additional specific standards and norms for the specific program evaluation about to be undertaken. This exercise helps surface any additional concerns and deepens buy-in to and ownership of the evaluation process. For example, one evaluation group I worked with established these norms for their evaluation:

- Strive for balance, looking at both the positive and negative.
- Respect confidentiality.
- Give the evaluation process a chance. Don't bad-mouth the evaluation and try to undermine it before we see what it yields.
- Be honest with the evaluator. Being honest, even about problems, is not disloyalty to the program.
- Collect both statistics and stories. This is part of balance.

This norm-setting exercise is especially useful with a group of particularly diverse stakeholders who come from different backgrounds, work in diverse professions with their own standards, and have diverse cultural experiences. This exercise offers them an opportunity to put forward to the group what they care about regarding how the evaluation is conducted. As the group deliberates about norms for and values that will guide the evaluation, they move from considering *whether to engage* genuinely with the evaluation process to *how to engage* genuinely with the process. In making that transition, they became ready to undertake a utilization-focused evaluation.

8. Sharing good ideas that haven't worked out in practice

This next exercise is one I like to assign groups at an evaluation launch session or capacity-building workshop just before taking a morning or afternoon break. I ask them to think of “a good idea you’ve had that didn’t work out in practice. It can be a work idea, something you tried in your personal life—anything at all. Something that seemed like a good idea at the time, but, well, things didn’t quite turn out as expected. And what evidence emerged that led you to realize that it hadn’t been such a good idea?” Then we take the planned 15-minute break and when they return, I ask them to share their examples in small groups. Then each small group picks one example to share with the full group. Often what they share is simple and funny. Nothing quite bonds a group together around a challenge like laughing together about life’s follies. Here are classic examples.

- “I thought it would be a good idea to buy the latest Microsoft computer operating system as soon as it came out. Crashed. Had all kinds of bugs. Took over my life doing updates and patches and rebooting. Now I wait until they work out the bugs.”
- “Trying to attract teenagers to a new program at our community center, we advertised free pizza. We had a great turnout and served the pizza right away to reward them for coming. But as soon as they had eaten, more than half left and didn’t stay for the program, didn’t even sign up for it. We learned to hold the pizza ’til near the end.”
- “As a public service during the presidential campaign, we decided to sponsor a debate with local reps of the two parties. We’re an employment training program, but don’t have anything to do with politics. We just wanted our participants to be informed and see democracy in action. Well, the two reps got very nasty with each other. Instead of discussing issues in a thoughtful way, they attacked each other’s motives and made derogatory comments about each other’s candidates. It was horrible. The debaters had each invited supporters to come and the evening denigrated into shouting. We were so naïve. We had done no screening of who was coming, set no ground rules, didn’t have a skilled moderator. We just thought, ‘It’s an election year. Let’s sponsor a little debate.’ Bad idea, at least the way we did it.”

These stories illustrate evaluative thinking at the most basic level. People had an idea about something they wanted to have happen (a goal). They tried it out (implementation). They observed what actually happened (data collection). Then made a judgment

(evaluation) and took away a lesson for the future (learning). Sharing these stories sets a context for the evaluation at hand. I follow up their stories with examples of good program ideas that didn't work out as hoped for, like the Kiribati overfishing initiative that opened the introductory chapter or the DARE example in which educating middle school students about drugs turned out not to work. They get the message: *We all have good ideas that don't work out in practice*. Utilization-focused evaluation isn't about blaming and criticizing and filling out forms and complying with funder mandates. It's about separating good ideas that work from seemingly good ideas that turn out not to work. It's about learning what works and doesn't work in order to do a better job in the future. As they absorb this message, they're getting ready to undertake their own program evaluation in a serious way—open to learning and change.

9. Generate evaluation metaphors and analogies

Key metaphors help determine what and how we perceive and how we think about our perceptions.

M. H. Abrams

Linguist and literary critic

Another way to ground evaluation in people's own experiences is to have them generate and share evaluation analogies and metaphors. For example, I'm based in Minnesota, where car license plates celebrate the *Land of 10,000 Lakes*—though the state actually has 11,842 lakes, a bit of a rounding error that, itself, raises the question of how accurate evaluation data need to be to be useful. But I digress. A lot of people in Minnesota fish—and those who don't are used to talking about fishing with those who do. So if I'm launching an evaluation in Minnesota, I'll often begin by asking: What constitutes a good day fishing? Well, it turns out that the answer depends on the purpose of fishing—whether it's with friends, or family; whether it's for food or recreation (catch and release), or some of each; and whether it's just for leisure or competitive, including high-stakes fishing contests. Different people have different criteria for what constitutes a good day fishing: number caught, size of fish caught, and type sought and reeled in, whether wall-eye, bluegill, northern pike, crappie, bass, or trout. Implicit in this example and discussion is that the seemingly straightforward question—What constitutes a good day fishing?—can be and will be answered differently by different people depending on their goals, values, criteria, preferences, the type of fishing they're doing, and who they're doing it with. And costs vary from very simple equipment (a pole, a line, a hook, and a lure) to very expensive fishing (a boat, a guide, fish-finding radar, high-tech poles, custom lures, and private access on exclusive lakes).

The same issues arise in program evaluation. The seemingly straightforward question—What constitutes a good program?—can be and will be answered differently by different

people depending on their goals, values, criteria, preferences, the type of program being implemented, the nature of the problem addressed, the types of participants targeted by the program, and the resources available for serving them. The purpose of bringing diverse stakeholders together for the evaluation launch is to discuss and come to agreement on these issues.

Any activity common to a group can serve as an analogy that grounds the evaluation discussion in shared experience. Sports, artistic endeavors, movie and theater preferences, vacation and holiday experiences, cooking, exercising, and major purchases (computers, cars) all involve goal-setting, criteria that distinguish good from bad and better from worse, differences of values and opinion, and data to inform decision making and judgment. Discussing such familiar activities and the evaluation issues embedded in them can be a warm-up exercise before turning to evaluation of the program at hand (like stretching before jogging).

Metaphors and analogies help us make connections between seemingly unconnected things and deepen our understanding of parallel issues. The language of research and evaluation—the jargon—is alien to many laypersons, decision makers, and stakeholders. From my point of view, the burden for clear communications rests on the evaluator. It is the evaluator who must find ways of bridging the communications gap. Metaphors and analogies help because:

Metaphor does more than adorn our thinking. It structures our thinking. It conditions our sympathies and emotional reactions. It helps us achieve situation awareness. It governs the evidence we consider salient and the outcomes we elect to pursue. (Klein, 1999, pp. 198–199)

When working with a group that I know by reputation or prior experience is relaxed together and open to a more playful and engaging process, I like to ask them to construct metaphors and similes about evaluation. The exercise helps participants discover their own values concerning evaluation while also giving them a mechanism to communicate those values to others. As with all the exercises in this chapter, it can be used with a program staff, an evaluation task force, evaluation trainees, workshop participants, or any group for whom it might be helpful to clarify and share perceptions about evaluation. The exercise opens like this.

One of the things that we'll need to do during the process of working together is come to some basic understandings about what evaluation is and can do. In my experience, evaluation can be a very creative and energizing experience. In particular, interpreting and using evaluation findings for program improvement requires creativity and openness to a variety of possibilities. To help us get started on this creative endeavor I'm going to ask you to participate with me in a little exercise.

In this box I have a bunch of toys, household articles, office supplies, tools, and other miscellaneous gadgets and thingamajigs that I've gathered from around my house. I'm going to dump these in the middle of the table and ask each of you to take one of them and use that item to make a statement about evaluation.

Evaluation is like _____ because . . .

To illustrate what I want people to do, I offer to go first. I ask someone to pick out any object in the room that I might use for my metaphor. What follows are some examples from actual workshops.

Someone points to a coffee cup: “This cup can be used to hold a variety of things. The actual contents of the cup will vary depending on who is using it and for what purpose they’re using it. Utilization-focused evaluation is a process like this cup; it provides a container but is empty until the group of people working on the evaluation fill it with focus and content and substance. The potential of the coffee cup cannot be realized until it holds some liquid. The potential of utilization-focused evaluation cannot be realized until it is given the substance of a concrete evaluation problem and situation. One of the things that I’ll be doing as we work together is providing an evaluation framework like this cup. You will provide the substance.”

Someone points to a chalkboard: “Evaluation is like a chalkboard because both are tools that can be used to express a variety of different things. The chalkboard itself is just an empty piece of slate until someone writes on it and provides information and meaning by filling in that space. The chalkboard can be filled up with meaningless figures, random marks, obscene words, mathematical formulas, or political graffiti—or the board can be filled with meaningful information, insights, helpful suggestions, and basic facts. The people who write on the chalkboard carry the responsibility for what it says. The people who fill in the blanks in the evaluation and determine its content and substance carry the responsibility for what the evaluation says. The evaluation process is just a tool to be used—and how it is used will depend on the people who control the process—in this case, you.”

I’ll typically take a break at this point and give people about 10 minutes to select an item and think about what to say. If there are more than 10 people in the group I will break the larger group into small groups of five or six for sharing analogies and metaphors so that each person is given an opportunity to make an evaluation statement. Below are some examples from actual workshops.

This empty grocery bag symbolizes my feelings about evaluation. When I think about our program being evaluated I want to find some place to hide. I can put this empty bag over my head so that nobody can see me and I can’t see anything else, and it gives me at least the feeling that I’m able to hide. (She puts the bag over her head.)

Evaluation can be like this toothbrush. When used properly it gets out the particles between the teeth so they don’t decay. If not used properly, if it just lightly goes over the teeth or doesn’t cover all the teeth, then some of the gunk will stay on and cause the teeth to decay. Evaluation should help get rid of any things that are causing a program to decay.

Evaluation is like this camera. It lets you take a picture of what's going on, but it can only capture what you point it at, and only for a particular point in time. My concern about this evaluation is that it won't give the whole picture, that an awful lot may get left out.

Evaluation for me is like this empty envelope. You can use it to send a message to someone. I want to use evaluation to send a message to our funders about what we're doing in the program. They don't have any idea about what we actually do. I just hope they'll read the letter when they get it.

Evaluation for me is like this adjustable wrench. You can use this wrench to tighten nuts and bolts to help hold things together. If used properly and applied with the right amount of pressure it holds things together very well. If you tighten the bolt too hard, however, you can break the bolt and the whole thing will fall apart. I'm in favor of evaluation if it's done right. My concern is that you can overdo it and the program can't handle it.

The process of sharing is usually accompanied by laughter and spontaneous elaborations of favorite metaphors. It's a fun process that offers hope the evaluation process itself may not be quite as painful as people thought it would be. In addition, participants are often surprised to find that they have something to say. They are typically quite pleased with themselves. Most important, the exercise serves to express very important thoughts and feelings that can be dealt with once they are made explicit.

Those participating are typically not even aware that they have these feelings. By providing a vehicle for discovering and expressing their concerns, it is possible to surface major issues that may later affect evaluation use. Shared metaphors can help establish a common framework for the evaluation, capturing its purpose, its possibilities, and the safeguards that need to be built into the process.

By the way, I've used this exercise with many different groups and in many different situations, including cross-cultural settings, and I've never yet encountered someone who couldn't find an object to use in saying something about evaluation. One way of guaranteeing this is to include in your box of items some things that have a pretty clear and simple message. For example, I'll always include a lock and key so that a very simple and fairly obvious analogy can be made: "Evaluation is like a lock and key, if you have the right key you can open up the lock and make it work. If you have the right information you can make the thing work." Or I'll include a lightbulb so that someone can say "evaluation is like this lightbulb, its purpose is to shed light on the situation."

10. Generating questions for evaluation

It is impossible for a man to learn what he thinks he already knows.

*Epictetus (CE 55-135),
Greek Stoic philosopher*

One way of facilitating a program's readiness for evaluation is to take primary intended users through a process of generating meaningful evaluation questions. I find that when

I enter a new program setting as an external evaluator, the people with whom I'm working typically expect me to tell them what the focus of the evaluation will be. They're passively waiting to be told by the evaluation expert—me—what questions the evaluation will answer. But I don't come with specific evaluation questions. I come with a process for determining what questions will be meaningful and what answers will be useful given the program's situation, priorities, and decision context. Taking them through the process of formulating questions and determining evaluation priorities is aimed at engendering their commitment to data-based evaluation and use.

This exercise involves getting stakeholders to think about what they want to learn from the evaluation without the jargon of or concern about goals, criteria, or measurement. On a chalkboard, flipchart, or handout I ask them to fill in the blank in the following sentence:

I WOULD REALLY LIKE TO KNOW _____
ABOUT THIS PROGRAM.

I ask each person, individually, to complete the blank 10 times. I ask: "What are 10 things about the program that you'd like to know, things you aren't certain about, that would make a difference in what you do if you had more information? Take a shot at it, without regard to methods, measurement, design, resources, precision—just 10 basic questions, real questions about this program."

After about 10 minutes I have them work in groups of four people each and ask them to combine their lists together into a single list of 10 things that *each group* wants to know, in effect, to establish each group's priority questions. Then we pull back together, the groups report, and together we generate a single list of 10 basic evaluation questions—answers to which could make a real difference to the effectiveness of the program.

The questions generated are usually the kind an experienced evaluator could anticipate. But being generated by the group, the questions are phrased in *their* terms, incorporating important local nuances of meaning and circumstance. Most important, they discover together that they have questions they care about—not my questions, as the external evaluator, but *their* questions, as people with some relationship to and stake in the program. In the process of generating, sharing, and prioritizing evaluation questions, it becomes *their* evaluation.

Later work together will involve refining questions, further prioritizing, formalizing evaluation procedures, determining methods, and establishing evaluation measures. But they aren't ready to engage in determining those details until they know they have questions that they want answered. This exercise increases their readiness for evaluation while allowing me, by seeing them in action and hearing what they come up with, assess their readiness

and determine what to do next in the evaluation process. Thus, generating a list of real and meaningful evaluation questions can play a critical part in getting key stakeholders involved in launching the evaluation process.

But generating a list of potentially useful questions is only one way to start interacting with primary users. How one begins depends on what backgrounds, experiences, preconceptions, and relationships the primary users bring to the table. If I need to get a group engaged quickly in reframing how they are thinking about my role as external evaluator, helping them discover their questions can move the focus from me to them and their stakes as stakeholders.



"I don't know what it is, but I feel like we're just not quite ready."

Initiating a Utilization-Focused Evaluation

Menu 1.1, pages 20–21, summarizes the 10 exercises we’ve just reviewed.

A utilization-focused evaluator begins by assessing program and organizational readiness for evaluation and using that assessment to begin strengthening the capacity to undertake a utilization-focused evaluation. Taking time at the beginning to understand and assess the program and organizational culture increases the likelihood that the evaluation can be designed to be relevant to and useful within the specific context within which it will unfold. Methods-driven evaluators often want to get right to work on designing the evaluation and generating instruments. Utilization-focused evaluators, in contrast, begin by learning about the context within which the evaluation will be conducted and, equally important, learning about the people within that context by getting them engaged in their own assessment of the evaluation context while deepening their commitment to serious reality testing.

Initial assessment and initial engagement go hand in glove. Assessing is a form of early engagement. Engaging stakeholders offers opportunities for both assessing and building the capacities of key stakeholders. The *utilization-focused checklist* in the book’s concluding chapter includes a summary of the key elements of Step 1. (See pp. 406–407.)

In closing, let me evoke the wisdom of Lao-tzu (604–531 BC), the Chinese philosopher who founded Taoism. He is famously attributed with the astute observation that “A journey of a thousand miles begins with a first step.” This chapter has been about that first step in utilization-focused evaluation.

However, this oft-cited quotation offers a deeper insight into the importance of understanding context when taking that first step. Michael Moncur (2004), a Chinese scholar, disputes this version of the popular quote and asserts that a more correct translation from the original Chinese would be, “The journey of a thousand miles begins beneath one’s feet.” He explains that rather than emphasizing the first step, Lao-tzu regarded action as something that arises naturally from stillness. Or, he goes on, another potential phrasing would be “Even the longest journey must begin where you stand.”

This chapter has emphasized finding out how key stakeholders understand, relate to, and translate EVALUATION. Different stakeholders will attach different meanings to the word, both cognitively and emotionally. They will also come with varying perceptions about what the evaluation journey will entail, many fearing that it will, indeed, be a journey of a thousand miles, or 10,000 miles, or even an interminable and quite treacherous journey. Assessing readiness to undertake the evaluation journey and deepening commitment to engage in the journey in a way that leads to useful results is where utilization-focused evaluation begins. It is an essential first step.

And I quite like the translation of Lao-tzu that says “Even the longest journey must begin where you stand.” We’ve been looking primarily at where the evaluation’s clients and primary stakeholders stand at the beginning. The next chapter looks at where the evaluator stands.



PRACTICE EXERCISES

1. Look up the full set of standards for evaluation: <http://www.jcsee.org/program-evaluation-standards>. Select one specific standard from each of the four overall categories (utility, feasibility, propriety, and accuracy) and explain how it would apply to a specific evaluation example.
2. Come up with a good idea of your own that didn't work out in practice. (See Exercise 8 in Menu 1.1.) What was the idea? How did you try to implement it? What was the "evidence" that led you to conclude it wasn't working? What did you learn from the experience? What does this example illustrate about evaluative thinking? How would you use this example with a group to help them get ready to engage in an evaluation?
3. Based on your own interests and expertise, develop a metaphor or analogy for evaluation. (See Exercise 9, Menu 1.1.) Explain how you would use it with a group.
4. The theme for the 2009 annual conference of the American Evaluation Association was *Context and Evaluation*. This chapter has been about how context matters and offered exercises to engage stakeholders in elucidating context: for example, assessing how past evaluations have been used and identifying incentives for and barriers to evaluation use.

Context typically refers to the setting (time and place) and broader environment in which the focus of the evaluation is located. Context also can refer to the historical context of the problem or phenomenon that the program or policy targets as well as the policy and decision-making context enveloping the evaluation. Context has multiple layers and is dynamic, changing over time (AEA, 2009).

Select a program that you are acquainted with and identify some contextual factors that would affect how an evaluation would be conducted.