7

Focusing on Outcomes

Beyond the Goals Clarification Game

Mulla Nasrudin was a Sufi guru. A king who enjoyed Nasrudin's company, and also liked to hunt, commanded him to accompany him on a bear hunt. Nasrudin was terrified.

When Nasrudin returned to his village, someone asked him: "How did the Hunt go?"

"Marvelously!"

"How many bears did you see?"

"None."

"How could it have gone marvelously, then?"

"When you are hunting bears, and you are me, seeing no bears at all is a marvelous experience."

—Shah 1964:61

Evaluation of the Bear Project

If this tale were updated by means of an evaluation report, it might read something like this:

Under the auspices of His Majesty's Ministry of the Interior, Department of Natural Resources, Section on Hunting, Office of Bears, field observers studied the relationship between the number of bears sighted on a hunt and the number of bears shot on a hunt. Having hypothesized a direct, linear relationship between the sighting of bears and killing of bears, data were collected on a recent royal hunting expedition.

The small sample size limits generalizability, but the results support the hypothesis at the 0.001 level of statistical significance. Indeed, the correlation is perfect. The number of bears sighted was zero and the number killed was zero. In no case was a bear killed without first being sighted. We therefore recommend new Royal regulations requiring that bears first be sighted before they are killed.
Respectfully submitted,

—The Incomparable Mulla Nasrudin
Royal Evaluator

Whose Goals Will Be Evaluated?

Although Nasrudin's evaluation bears (and bares) certain flaws, it shares one major trait with almost all other reports of this genre: It is impossible to tell whether it answers anyone's question. Who decided that the goal evaluated should be the number of bears killed? Perhaps the hunt's purpose was a heightened sensitivity to nature, or a closer relationship between Nasrudin and the king, or reducing Nasrudin's fear of bears, or increasing the king's power over Nasrudin. It may even be possible (likely!) that different participants in the hunt had different goals. Nasrudin perceived a "marvelous" outcome. Other stakeholders, with different goals (e.g., bagging a bear), might have concluded otherwise.

In utilization-focused evaluation, the primary intended users determine whose goals will be evaluated if they decide that evaluating goal attainment will be the focus of the evaluation. There are other ways of focusing an evaluation, as we'll see in the next chapter, but first, let's review the traditional centrality of goal attainment in evaluation.

The Centrality of Goals in Evaluation

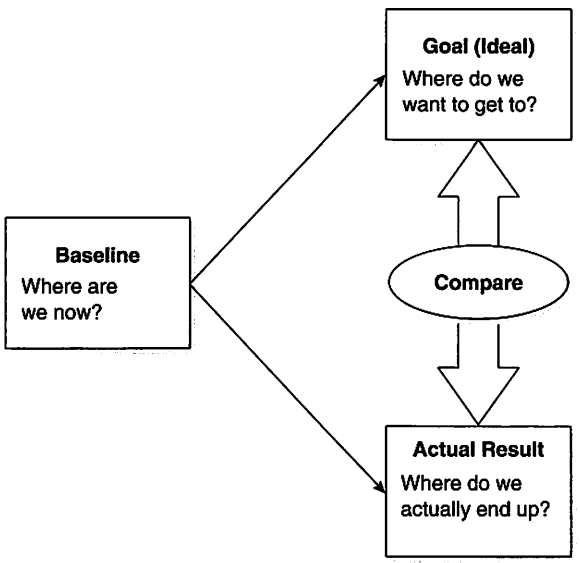
Traditionally, evaluation has been synonymous with measuring goal attainment (Morris and Fitz-Gibbon 1978). The basic logic of goals-based evaluation involves, at a minimum, three points of comparison: (1) a starting point, or baseline; (2) a goal or target (the ideal); and (3) the ending point, or actual result. It doesn't get any simpler than that. Exhibit 7.1 depicts this fundamental ideal-actual goals-based logical comparison.

Distinguished evaluation methodologist Peter Rossi (1972) asserted that "a social welfare program (or for that matter any program) which does not have clearly specified goals cannot be evaluated without specifying some measurable goals. This statement is obvious enough to be a truism" (p. 18). In a major review of the evaluation literature in education, Worthen and Sanders (1973) concluded that "if evaluators agree on anything, it is that program objectives written in unambiguous terms are useful information for any evaluation study" (p. 231). Carol Weiss (1972b) observed that

> the traditional formulation of the evaluation question is: To what extent is the program succeeding in reaching its goals?. . . . The goal must be clear so that the evaluator knows what to look for. . . . Thus begins the long, often painful process of getting people to state goals in terms that are *clear*, *specific*, and *measurable*. (Pp. 74–76)

EXHIBIT 7.1

Most Fundamental Goals-Based Logic of Evaluation: Ideal-Actual Comparison



Stating Goals Can Be Tricky

A Youth Center highlighted the following goal in its funding proposal to a philanthropic foundation:

> We strive hard for every participant to achieve their goals and dreams by nurturing their self esteem as they grow up. It is our goal that 85% will feel motivated or good about themselves and 15% will not.

As the preceding quotes illustrate, the evaluation literature is replete with serious entreaties on the centrality of program goals, and this solemnity seems to carry over into evaluators' work with program staff. There may be no more deadly way to begin an evaluation effort than assembling program staff to identify and clarify program goals and objectives. If evaluators are second only to tax collectors in the hearts of program staff, I suspect that it is not because staff fear evaluators' judgments about program success but because they hate constant questioning about goals.

The Goals Clarification Game

Evaluators frequently conduct goals clarification meetings as if playing the Twenty Questions party game. Someone thinks of an object in the room and then the players are allowed 20 questions to guess what it is. In the goals clarification game, the evaluator has an object in mind (a clear, specific, and measurable goal). Program staff are the players. The game begins with the staff generating some statement they think is a goal. The evaluator scrutinizes the statement for clarity, specificity, and measurability, usually judging the staff's effort inadequate. This process is repeated in successive tries until the game ends in one of three ways: (1) The staff gives up (so the evaluator wins and writes the program goals for staff); (2) the evaluator gives up (so the staff gets by with vague, fuzzy, and unmeasurable goals); or (3) in rare cases, the game ends when staff actually stumbles on a statement that reasonably approximates what the evaluator had in mind.

Why do program staff typically hate this game so much?

1. They have played the game hundreds of times, not just for evaluators, but for funders and advisory boards, in writing proposals, and even among themselves.

2. They have learned that when playing the game with an evaluator, the evaluator almost always wins.

3. They come out of the game knowing that they appear fuzzy-minded and inept to the evaluator.

4. It is a boring game.

5. It is an endless game because each new evaluator comes to the game with a different object in mind. (Clarity, specificity, and measurability are not clear, specific, and measurable criteria, so each evaluator can apply a different set of rules in the game.)

Among experienced program staff, evaluators may run into countering strategies like the goals clarification shuffle. Like many dance steps (e.g., the Harlem shuffle, the hustle), this technique has the most grace and style when executed simultaneously by a group. The goals clarification shuffle involves a sudden change in goals and priorities after the evaluator has developed measuring instruments and a research design. The choreography is dazzling. The top-priority program goal is moved two spaces to either the right or left and four spaces backward. Concurrently, all other goals are shuffled with style and subtlety, the only stipulation being that the first goal end up somewhere in the middle, with other goals reordered by new criteria.

The goals clarification shuffle first came into national prominence in 1969 when it was employed as a daring counterthrust to the Westinghouse-Ohio State University Head Start Evaluation. That study evaluated cognitive and affective outcomes of the Head Start Program and concluded that Head Start was largely ineffective (Cicarelli 1971; Westinghouse Learning Corporation 1969). However, as soon as the final report was published, the goals clarification shuffle was executed before enthusiastic Congressional audiences, showing that Head Start's health, nutrition, resource redistribution, cultural, and community goals ought to have been in the spotlight (see Evans 1971:402; Williams and Evans 1969). Thus, despite negative evaluation findings, Congress expanded the Head Start program, and the evaluators were thrown on the defensive. (It was about this same time that serious concerns over nonuse of evaluation findings started to be heard on a national scale.)

Conflicts Over Goals

Not all the goals clarification exercises resemble dances. Often, the more fitting metaphor is competition. Conflict over program goals among different stakeholder groups is common. For example, in criminal justice programs, battles are waged over whether the purpose of a program is punitive (punish criminal offenders for wrongdoing), custodial (keep criminal offenders off the streets), or rehabilitative (return offenders to society after treatment). In education and training programs, conflicts often emerge over whether the priority goal is attitude change or behavior change. In welfare agencies, disagreements can be found over whether the primary purpose is to get clients off welfare or out of poverty, and whether the focus should be long-term change or short-term crisis intervention (Conte 1996). In health settings, staff dissension may emerge over the relative emphasis to be placed on preventive versus curative medical practice. Chemical dependency programs are often enmeshed in controversy over whether the desired outcome is sobriety or responsible use. Even police and fire departments can get caught in controversy about the purpose and actual effects of sirens, with critics arguing that they're more a nuisance than a help (Perlman 1996). Virtually any time a group of people assemble to determine program goals, conflict can emerge, resulting in a lengthy, frustrating, and inconclusive meeting.

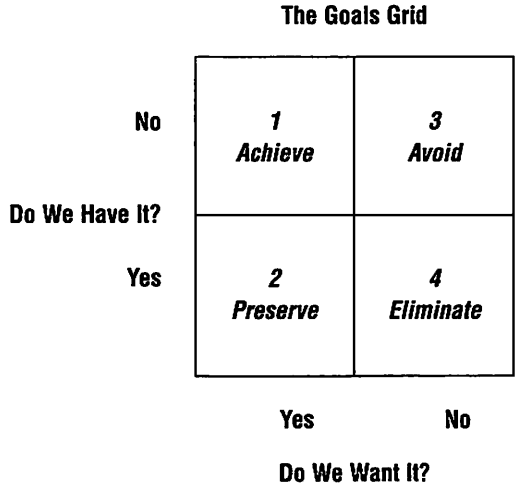
For inexperienced evaluators, conflicts among stakeholders can be unnerving. Once, early in my career, a goals clarification session erupted into physical violence between a school board member and the district's internal evaluator. The novice evaluator can lose credibility by joining one side or the other. More experienced evaluators have learned to remain calm

EXHIBIT 7.2

Conceptualizing goals involves identifying a gap between what people have and what they want. This gap or discrepancy can take several forms. Fred Nickols (2003) developed *The Goals Grid* to help people clarify goals. In some cases, people want things they don't have and in others they have things they don't want. This generates two questions: "Do we have it?" and "Do we want it?" The four goal categories then become

1. Achieve (we don't have it and we want it)
2. Preserve (we have it and we want to keep it)
3. Avoid (we don't have it and we don't want it)
4. Eliminate (we have it and we don't want it)

This Goals Grid is a useful tool for thinking about different kinds of program and participant goals and objective.



SOURCE: Reprinted with permission of Fred Nickols.

and neutral, sometimes suggesting that multiple goals be evaluated, thereby finessing the need for consensus about program priorities.

In some instances, an evaluator may encounter intense fighting over goals and values. A "goals war" usually occurs when two or more strong coalitions are locked in battle to determine which group will control the future direction of some public policy or program. Such conflicts involve highly emotional issues and deeply held values, such as conflicting views on abortion or sex education for teenagers.

Evaluation of school busing programs to achieve racial balance offers an example rich with conflict. By what criteria ought busing programs be evaluated? Changed

racial attitudes? Changed interracial behaviors? Improved student achievement? Degree of parent involvement? Access to educational resources? All are candidates for the honor of primary program goal. Is school busing supposed to achieve desegregation (representative proportions of minority students in all schools) or integration (positive interracial attitudes, cooperation, and interaction)? Many communities, school boards, and school staffs are in open conflict over these issues. Central to the battles fought are basic disagreements about what evaluation criteria to apply.

Classic Goals Conflicts

Prisons:	rehabilitation versus punishment
Chemical dependency:	sobriety versus responsible drinking
Homelessness:	adequate shelters versus low-income housing
Sex education:	abstinence versus responsible safe sex
Affirmative action:	equal opportunity versus equal results
Welfare reform:	off welfare versus out of poverty
Transportation:	more freeways for cars versus public transportation
Abortion:	restrictions on choice versus informed choice and access
Health care:	treatment versus prevention
School desegregation:	racial integration versus equal achievement

Evaluability Assessment and Goals Clarification

Evaluators have gotten heavily involved in goals clarification because, when we are invited in, we seldom find a statement of clear, specific, prioritized, and measurable goals. This can take novice evaluators by surprise if they think that their primary task will be formulating an evaluation design for already established goals. Even where goals exist, they are frequently unrealistic, having been exaggerated to secure funding—what are called *BHAG*s (big hairy audacious goals). One reason evaluability assessment has become an important preevaluation tool is that by helping programs get ready for evaluation, it acknowledges the frequent need for a period of time to work with program staff, administrators, funders, and participants on clarifying goals—making them realistic, meaningful, agreed on, and evaluable (Wholey 1994; Smith 1989). Evaluability assessment often includes fieldwork and interviews to determine how much consensus there is among various stakeholders about goals and to identify where differences lie. Based on this kind of contextual analysis, an evaluator can work with primary intended users to plan a strategy for goals clarification.

When an evaluability assessment reveals broad aims and fuzzy goals, it's important to understand what role goals are understood to play in the program. Fuzzy goals actually characterize much human cognition and reasoning (Ragin 2000; Zadeh et al. 1975). Classic laboratory experiments found that fuzzy conceptualizing may be typical of half the population (Kochen 1975:407). No wonder evaluators have so much trouble getting clear, specific, and measurable goals! Carol Weiss (1972b) has commented in this regard:

> Part of the explanation [for fuzzy goals] probably lies in practitioners' concentration on concrete matters of program functioning and their pragmatic mode of operation. They often have an intuitive rather than an analytical approach to program development. But there is also a sense in which ambiguity serves a useful function; it may mask underlying divergences in intent . . . glittering generalities that pass for goal statements are meant to satisfy a variety of interests and perspectives. (P. 27)

Thus, evaluators have to figure out if administrators and staff are genuinely fuzzy about what they're attempting to accomplish, or if they're simply being shrewd in not letting the evaluator (or others) discover their *real* goals, or if they're trying to avoid conflict through vagueness. Fuzzy goals, then, may be a conscious strategy for avoiding conflict among competing interests. In such instances, the evaluation may be focused on important questions, issues, and concerns without resort to clear, specific, and measurable objectives. However, more often than not in my experience, the difficulty turns out to be a conceptual problem rather than deviousness.

From a utilization-focused point of view, the challenge is to calculate how early interactions in the evaluation process will affect later use. Typically, it's not useful to ignore goals conflict, accept poorly formulated or unrealistic goals, or let the evaluator assume responsibility for writing clear, specific, and measurable goals. Primary intended users need to be involved in assessing how much effort to put into goals clarification. In doing so, both evaluators and primary intended users do well to heed the *evaluation standard on political viability:*

> The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to

> bias or misapply the results can be averted or counteracted. (Joint Committee on Standards 1994:F2)

There are alternatives to goals-based evaluation, alternatives we'll consider in the next chapter. First, let's examine how to work with intended users who want to focus on goals and results.

Communicating about Goals and Results

Part of the difficulty, I am convinced, is the terminology *goals and objectives.* These very words can intimidate staff. Goals and objectives have become daunting weights that program staff feel around their necks, burdening them, slowing their efforts, and impeding rather than advancing their progress. Helping staff clarify their purpose and direction may mean avoiding use of the term goals and objectives.

I've found program staff quite animated and responsive to the following kinds of questions: What are you trying to achieve with your clients? If you are successful, how will your clients be different after the program than they were before? What kinds of changes do you want to see in your clients? When your program works as you want it to, how do clients *behave* differently? What do they say differently? What would I see in them that would tell me they are different? Program staff can often provide quite specific answers to these questions, answers that reveal their caring and involvement with the client change process, yet when the same staff are asked to specify their goals and objectives, they freeze.

Questions for Clarifying Goals and Intended Outcomes

What are you trying to achieve with your program participants?

If you are successful, how will participants be different after the program than they were before?

What kinds of changes do you want to see in program participants? When your program works as you want it to, how do participants *behave* differently?

What do they say differently? What would I see in them that would tell me they are different?

After querying staff about what results they hope to accomplish with program participants, I may then tell them that what they have been telling me constitutes their goals and objectives. This revelation often brings considerable surprise. They often react by saying, "But we haven't said anything about what we would count." This, as clearly as anything, I take as evidence of how widespread the confusion is between the conceptualization of goals and their measurement. Help program staff and other intended users be realistic and concrete about goals and objectives, but don't make them hide what they are really trying to do because they're not sure how to write a formally acceptable statement of goals and objectives, or because they don't know what measurement instruments might be available to get at some of the important things they are trying to do. Instead, *take them through a process that focuses on achieving outcomes and results rather than writing goals.* The difference, it turns out, can be huge.

Focusing on Outcomes and Results

In the minds of many program people, from board members to frontline staff and participants, goals are abstract statements of ideals written to secure finding—meant to inspire but never achieved. Consider this poster on the wall of the office of a program I evaluated:

> *The greatest danger is not that we aim too high and miss, but that our goal is too low and we attain it.*

For the director of this program, goals were something you put in proposals and plans, and hung on the wall, then went about your business.

Let me illustrate the difference between traditional program goals and a focus on participant outcomes with plans submitted by county units to a state human services agency. The plans required statements of outcomes. Each statement below promises something, but that something is not a change in client functioning, status, or well-being. These statements reveal how people in social services have been trained to think about program goals. My comments, following each goal, are meant to illustrate how to help program leaders and other intended evaluation users reframe traditional goals to focus on participant outcomes.

Problematic Outcome Examples

1. To continue implementation of a case management system to maintain continued contact with clients before, during, and after treatment.

Comment: Continued implementation of the system is the goal. And what is promised for the client? "Continued contact."

2. Case management services will be available to all persons with serious and persistent mental illness who require them.

Comment: This statement aims at availability—a service delivery improvement. Easily accessible services could be available 24 hours a day, but with what outcomes?

3. To develop needed services for chronically chemically dependent clients.

Comment: This statement focuses on program services rather than the client outcomes. My review of county plans revealed that most managers focus planning at the program delivery level, that is, the program's goals, rather than how clients' lives will be improved.

4. To develop a responsive, comprehensive crisis intervention plan.

Comment: A plan is the intended outcome. I found that many service providers confuse planning with getting something done. The characteristics of the plan—"responsive, comprehensive"—reveal nothing about results for intended beneficiaries.

5. Develop a supportive, family-centered empowering, capacity-building intervention system for families and children.

Comment: This goal statement has lots of human services jargon, but, carefully examined, the statement doesn't commit to empowering any families or actually enhancing the capacity of anyone receiving services through the system.

6. Expand placement alternatives.

Comment: More alternatives is the intended results, but to what end? Here is another system-level goal that carries the danger of making placement an end in itself rather than a means to client improvement.

7. County clients will receive services that they value as appropriate to their needs and helpful in remediating their concerns.

Comment: Client satisfaction can be an important outcome, but it's rarely sufficient by itself. Especially in tax-supported programs, taxpayers and policymakers

want more than happy clients. They want clients to have jobs, be productive, stay sober, parent effectively, and so on. Client satisfaction needs to be connected to other desired outcomes.

8. Improve ability of adults with severe and persistent mental illness to obtain employment.

Comment: Some clients remain for years in programs that enhance their ability to obtain employment—without ever getting a job.

9. Adults with serious and persistent mental illness will engage in a process to function effectively in the community.

Comment: Engaging in the process is as much as this aims for, in contrast to clients actually functioning effectively in the community.

10. Adults with developmental disabilities will participate in programs to begin making decisions and exercising choice.

Comment: Program participation is the stated focus. This leads to counting how many people show up rather than how many make meaningful decisions and exercise real choice. A client can participate in a program aimed at teaching decision-making skills, and can even learn those skills, yet never be permitted to make real decisions.

11. Each developmentally disabled consumer (or their substitute decision maker) will identify ways to assist them to remain connected, maintain, or develop natural supports.

Comment: This goal is satisfied, as written, if each client has a list of potential connections. The provider, of course, can pretty much guarantee composition of such a list. The actual important outcome: Clients who are connected in a meaningful way to a support group of people.

12. Adults in training and rehab will be involved in an average of 120 hours of community integration activities per quarter.

Comment: Quantitative and specific, but the outcome stated goes only as far as being involved in activities, not actually being integrated into the community.

13. Key indicators of intended results and client outcomes for crisis services:

- Number of patients served
- Number of patient days and the average length of stay
- Source of referrals to the crisis unit and referrals provided to patient at discharge

Comment: Participation numbers, not client outcomes.

14. Minimize hospitalizations of people with severe and persistent mental illness.

Comment: This is a system-level outcome that is potentially dangerous. One of the premises of results-oriented management is that "what gets measured gets done." An easy way to attain this desired outcome is simply not to refer or admit needy clients to the hospital. That will minimize hospitalizations (a system-level outcome) but may not help people in need. A more appropriate outcome would be that these clients function effectively. If that outcome is attained, they won't need hospitalizations.

15. Improve quality of child protection intervention services.

Comment: I found a lot of outcome statements aimed at enhancing quality. Ironically, quality can be enhanced by improving services without having an impact on client outcomes. Licensing and accrediting standards often focus on staff qualifications and site characteristics (indicators of quality), but seldom require review of what program participants achieve.

The point of reviewing these examples has been to show the kinds of goal statements an evaluator may encounter when beginning to work with a program. A utilization-focused evaluator can help intended users review plans and stated goals to see if they include an outcome focus. There's nothing wrong with program-level goals (e.g., improve access or quality) or system-level goals (e.g., reduce costs), but such goals ought to connect to outcomes for clients. An evaluator can facilitate discussion of why, in the current political environment, one hears increased demand for "outcomes-based" management and accountability (e.g., Nolan and Mock 2005). Given that increased focus, there are helpful guides for working with outcomes in evaluation, such as the excellent GrantCraft guide *Making Measures Work for You: Outcomes and Evaluation*, supported by the Ford Foundation (McGarvey 2006), the Bruner Foundation capacity-building resources (Baker 2008), and *Getting to Outcomes with Developmental Assets* (Fisher, Chinman, and Wandersman 2006). Evaluators need to provide technical assistance in helping program planners, managers, and other potential evaluation users understand the difference between a participant outcomes approach and traditional program or system goals approaches. In particular, they need assistance understanding the difference between service-focused goals versus client-focused outcome goals. Exhibit 7.3 compares these two kinds of goals. Both can be useful, but they place emphasis in different places.

EXHIBIT 7.3

Service-Focused Versus Outcome-Focused Goals: Examples From Parenting Programs

Service-Focused Goals	*Outcome-Focused Goals*
Provide coordinated case management services with public health to pregnant adolescents	Pregnant adolescents will give birth to healthy babies and care for the infants and themselves appropriately
Improve the quality of child protection intervention services	Children will be safe; they will not be abused or neglected
Develop a supportive, family-centered, capacity-building intervention system for families and children	Parents will adequately nurture and provide necessities for their children
Provide assistance to parents to make employment-related child care decisions	Parents who wish to work will have adequate child care

Leading a Horse to Water versus Getting It to Drink

The shift from service goals to outcomes often proves difficult in programs and agencies that have a long history of focusing on services and activities. But even where the difference is understood and appreciated, some fear or resistance may emerge. One reason is that service providers are well schooled in the proverbial wisdom that "you can lead a horse to water, but you can't make it drink."

This familiar adage illuminates the challenge of committing to outcomes. The desired outcome is that the horse drink the water. Longer-term outcomes are that the horse stays healthy and works effectively. But because program staff know they can't make a horse drink water, they focus on the things they can control: leading the horse to water, making sure the tank is full, monitoring the quality of the water, and keeping the horse within drinking distance of the water. In short, they focus on the *processes* of water delivery rather than the outcome of water drunk. Because staff can control processes but cannot guarantee attaining outcomes, government rules and regulations get written specifying exactly how to lead a horse to water. Funding is based on the number of horses led to water. Licenses are issued to individuals and programs that meet the qualifications for leading horses to water. Quality awards are made for improving the path to the water—and keeping the horse happy along the way. Whether the horse drinks the water gets lost in all this flurry of lead-to-water-ship. Most reporting systems focus on how many horses get led to the water, and how difficult it was to get them there, but never quite get around to finding out whether the horses drank the water and stayed healthy. And they seem unaware of the parallel wisdom that "you may have to lead a horse to water but it will run on its own to oats." Or this, sometimes called Murphy's 12th law: "You can't lead a cavalry charge if you think you look funny on a horse." But enough with horses already. Back to evaluation and accountability.

One point of resistance to outcomes accountability, then, is the fear among providers and practitioners that they're being asked to take responsibility for, and will be judged on, something over which they have little control. The antidote to this fear is involving staff in determining how to measure outcomes and establishing a results-oriented culture in an organization or agency. Evaluators have a role to play in such efforts by facilitating a process that helps staff, administrators, and other stakeholders think about, discuss the implications of, and come to understand both the advantages and limitations of an outcomes approach. There's a lot of managerial and political rhetoric about being results oriented, but not much expertise in how to set up a results-oriented system. The next section presents a framework for conceptualizing outcomes that are meaningful and measurable for use in facilitating an outcomes-oriented management, monitoring, and evaluation system.

Utilization-Focused Outcomes Framework

This framework distinguishes six separate elements that need to be specified for focusing an evaluation on participant or client outcomes:

- A specific participant or client target group
- The desired outcome(s) for that target group
- One or more indicators for each desired outcome
- Details of data collection
- How results will be used
- Performance targets

I'll discuss each of these elements and offer illustrations from actual programs to show how they fit together. Evaluators can use this framework to work with primary intended users.

Identifying Specific Participant or Client Target Groups

I'll use the generic term *client* to include program participants, consumers of services, beneficiaries, students, and customers, as well as traditional client groups. The appropriate language varies, but for every program, there is some group that is expected to benefit from and attain outcomes as a result of program participation. However, the target groups identified in enabling legislation or existing reporting systems typically are defined too broadly for meaningful outcomes measurement. Intended outcomes can vary substantially for subgroups within general eligible populations. The trick is to be as specific as necessary to conceptualize meaningful outcomes. Some illustrations may help clarify why this is so.

Consider a program aimed at supporting the elderly to continue living in their homes, with services ranging from "meals on wheels" to home nursing. Not all elderly people can or want to stay in their homes. Therefore, if the desired outcome is "continuing to live in their own home," it would be inappropriate to specify that outcome for all elderly people. A more appropriate target population, then, would be people above the age of 65 who want to and can remain safely in their homes. For this group, it is appropriate to aim to keep them in their homes. It is also clear that some kind of screening process will be necessary to identify this subpopulation of the elderly.

A different example comes from programs serving people with developmental disabilities (DD). Many programs exist to prepare DD clients for work and then support them in maintaining employment. However, not all people with DD can or want to work. In cases where funding supports the right of DD clients to choose whether to work, the appropriate subpopulation becomes people with DD who can and want to work. For that specific subpopulation, then, the intended outcome could be that they obtain and maintain satisfying employment.

There are many ways of specifying subpopulation targets. Outcomes are often different for young, middle-aged, and elderly clients in the same general group (e.g., persons with serious and persistent mental illness). Outcomes for pregnant teens or teenage mothers may be different from outcomes for mothers receiving welfare who have completed high school. Outcomes for first-time offenders may be different from those for repeat offenders. The point is that categories of funding eligibility often include subgroups for which outcomes appropriately vary. Similarly, when identifying groups by services received, for example, counseling services or jobs training, the outcomes expected for generic services may vary by subgroups. It is important, then, to make sure an intended outcome is meaningful and appropriate for everyone in the identified target population.

Specifying Desired Outcomes

The choice of language varies under different evaluation approaches. Some models refer to *expected outcomes* or *intended outcomes*. Others prefer the language of *client goals* or *client objectives*. What is important is not the phrase used but that there be a clear statement of the targeted change in circumstances, status, level of functioning, behavior, attitude, knowledge, or skills. Other outcome types include maintenance and prevention. Exhibit 7.4 provides examples of outcomes.

EXHIBIT 7.4
Outcome Examples

Type of Change	*Illustration*
Change in circumstances	Children safely reunited with their families of origin from foster care
Change in status	Unemployed to employed
Change in behavior	Truants will regularly attend school
Change in functioning	Increased self-care; getting to work on time
Change in attitude	Greater self-respect
Change in knowledge	Understand the needs and capabilities of children at different ages
Change in skills	Increased reading level; able to parent appropriately
Maintenance	Continue to live safely at home (e.g., the elderly)
Prevention	Teenagers will not use drugs

Outcome Indicators: Operationalizing

An indicator is just that, an indicator. It's not the same as the phenomenon of interest but only an indicator of that phenomenon. A score on a reading test is an indicator for reading capability but should not be confused with a particular person's true ability. All kinds of things affect a test score on a given day. Thus, indicators are inevitably approximations. They are imperfect and vary in validity and reliability.

Figuring out how to measure a desired outcome is called *operationalizing* the outcome. In selecting a measurement, the desired outcome becomes more than a concept; it becomes something that can be operated. Indeed, the common dictionary definition of operational is that it makes it "*ready for use*" (American Heritage Dictionary 2006). A desired outcome isn't ready for use in evaluation until it is operationalized.

The resources available for measurement will greatly affect the kinds of data that can be collected for indicators. For example, if the desired outcome for abused children is that there be no subsequent abuse or neglect, regular in-home visitations and observations, including interviews with the child, parent(s), and knowledgeable others, would be desirable, but such data collection is expensive. With constrained resources, one may have to rely on routinely collected data and mandated reporting, that is, official substantiated reports of abuse and neglect over time. Moreover, when using such routine data, privacy and confidentiality restrictions may limit the indicator to

aggregate results quarter by quarter rather than one that tracks specific families over time. In contemporary child protection services, tracking substantiated abuse, while desirable and required, is not considered enough and systems that enable much more detailed data collection have been and are being used. Risk assessment protocols not only attempt to establish risk of re-abuse but actually may be used to determine the frequency of agency contact with families.

As resources change, the indicator may change. Routine statistics may be used by an agency until a philanthropic foundation funds a focused evaluation to get better data for a specific period of time. In such a case, the indicator would change, but the desired outcome would not. This is the advantage of clearly distinguishing the desired outcome from its indicator. *As the state of the art of measurement develops or resources change, indicators may improve without changing the desired outcome.*

Time frames also affect indicators. The ultimate goal of a program for abused children would be to have them become healthy, well-functioning, and happy adults, but policymakers cannot wait 10 to 15 years to assess the outcomes of a program for abused children. Short-term indicators must be relied on, things such as school attendance, school performance, physical health, and the psychological functioning of a child, as well as any redeterminations of abuse. These short-term indicators provide sufficient information to make judgments about the likely long-term results. It takes 30 years for a forest to grow, but you can assess the likelihood of ending up with a forest by evaluating how many saplings are still alive, 1 year and 3 years after the trees are planted.

Another factor affecting indicator selection is the demands data collection will put on program staff and participants. Short-term interventions such as food shelves, recreational activities for people with developmental disabilities, drop-in centers, and one-time community events do not typically engage participants intensely enough to justify collection of much, if any, data. Many programs can barely collect data on end-of-program status, much less follow-up data 6 months after program participation.

In short, a variety of factors influence the selection of indicators, including the importance of the outcome claims being made, resources available for data collection, the state of the art of measurement of human functioning, the nature of decisions to be made with the results, and the willingness of staff and participants to engage in assessment. Some kind of indicator is necessary, however, to measure degree of outcome attainment. *The key is to make sure that the indicator is a reasonable, useful, and meaningful measure of the intended client outcome.*

The framework offered here will generate outcome statements that are *clear, specific,* and *measurable,* but getting clarity and specificity is separated from selecting measures. The reason for separating the identification of a desired outcome from its measurement is to ensure the utility of both. This point is worth elaborating. The following is a classic goal statement:

> Student achievement test scores in reading will increase one grade level from the beginning of first grade to the beginning of second grade.

Such a statement mixes together and potentially confuses the (1) specification of a desired outcome with (2) its measurement

and (3) the desired performance target. The desired outcome is increased student achievement. The indicator is the score on a norm-referenced standardized achievement test. The performance target is one year's academic gain on the test. These are three separate decisions that primary intended evaluation users need to discuss. For example, there are ways other than standardized tests for measuring achievement, such as student portfolios or competency-based tests. The desired outcome should not be confused with its indicator. In the framework offered here, outcome statements are clearly separated from operational criteria for measuring them.

Another advantage of separating outcomes identification from indicator selection is to encourage program staff to be serious about the process. A premature focus on indicators may be heard as limiting a program to attempt only those things that staff already know how to measure. Such a limitation is too constraining. It is one thing to establish a purpose and direction for a program. It is quite another thing to say how that purpose and direction are to be measured. By confusing these two steps and making them one, program goals can become detached from what program staff and funders are actually working to accomplish. Under such a constraint, staff begin by figuring out what can be measured. Given that they seldom have much expertise in measurement, they end up counting fairly insignificant behaviors and attitudes that they can somehow quantify.

When I work with groups on goals clarification, I have them state intended outcomes *without regard to measurement.* Once they have stated as carefully and explicitly as they can what they want to accomplish, then it is time to figure out what indicators and data can be collected to monitor outcome attainment. They can then move back and forth between conceptual level statements and operational (measurement) specifications, attempting to get as much precision as possible in both.

To emphasize this point, let me overstate the trade-off. *I prefer to have less-than-ideal or rough measures of important goals rather than highly precise measures of goals that no one much cares about.* In too many cases, evaluators focus on the later (meaningless but measurable goals) instead of on the former (meaningful goals with less-than-ideal measures).

Of course, this trade-off, stated in stark terms, is only relative. It is desirable to have as much precision as possible. By separating the process of goals clarification from the process of selecting indicators, it is possible for program staff to focus first on what they are really trying to accomplish and to state their goals and objectives as explicitly as possible *without regard to measurement*, and then to worry about how one would measure actual attainment of those goals and objectives.

SMART Indicators

SMART is a widely used mnemonic for helping people remember the characteristics of a good indicator.

Specific

Measurable

Achievable

Relevant

Timebound

Performance Targets

A performance target specifies the amount or level of outcome attainment that is expected, hoped for, or, in some kinds of performance contracting, required. What percentage of participants in employment training will have full-time jobs 6 months after graduation: 40 percent? 65 percent? 80 percent? What percentage of fathers failing to make child support payments will be meeting their full child support obligations within 6 months of intervention? 15 percent? 35 percent? 60 percent?

The best basis for establishing future performance targets is past performance. "Last year we had 65 percent success. Next year we aim for 70 percent." Lacking data on past performance, it may be advisable to wait until baseline data have been gathered before specifying a performance target. Arbitrarily setting performance targets without some empirical baseline may create artificial expectations that turn out unrealistically high or embarrassingly low. One way to avoid arbitrariness is to seek norms for reasonable levels of attainment from other, comparable programs, or review the evaluation literature for parallels.

As indicators are collected and examined over time, from quarter to quarter, and year to year, it becomes more meaningful and useful to set performance targets. The relationship between resources and outcomes can also be more precisely correlated longitudinally, with trend data, all of which increases the incremental and long-term value of an outcomes management approach.

The challenge is to make performance targets meaningful. Chapter 13, on interpreting data, includes further discussion on performance targets and establishing standards of desirability that can be developed with primary intended users as a framework for interpreting outcomes data.

In a political environment of outcomes mania, meaningfulness and utility are not necessarily priorities. Consider this example and judge for yourself. The 1995 Annual Management Report from the Office of the New York City Mayor included this performance target: The average daytime speed of cars crossing from one side of midtown Manhattan to the other will increase from 5.3 to 5.9 miles per hour. Impressed by this vision of moving from a "brisk 5.3" to a "sizzling 5.9," *The New Yorker* magazine interviewed Ruben Ramirez, Manhattan's Department of Transportation Traffic Coordinator, to ask how such a feat could be accomplished in the face of downsizing and budget cuts. Ramirez cited better use of resources. Asked what could he accomplish with adequate resources, he replied, "I think we could do six or seven, and I'm not being outrageous." *The New Yorker* found such a performance target a "dreamy future," one in which it might actually be possible to drive across midtown Manhattan faster than you can walk ("Speed" 1995:40).

Is such a vision visionary? Is a performance increase from 5.3 to 5.9 miles per hour meaningful? Is 6 or 7 worth aiming for? For a noncommuting Minnesotan, such numbers fail to impress. But, converted into annual hours and dollars saved for commercial vehicles in Manhattan, the increase may be valued in hundreds of thousands of dollars, perhaps even millions. It's for primary stakeholders in Manhattan, not Minnesota, to determine the meaningfulness of such a performance target.

Trade-Offs in Setting Performance Targets

In 2007, the United States Internal Revenue Service (IRS) began auditing income tax returns less deeply and more broadly, amounting to a change in its performance targets. Formerly the IRS target was collecting "every last dime" in an audit to maximize the amount collected in each individual audit. The new performance target is collecting 80 percent of what might be available in an audit but, at the same time, conducting a greater number of audits.

What's the rationale for this change? It involves a trade-off based on yet another version of an 80/20 rule. The IRS has found that its agents can collect 80 percent of the tax recoverable in an audit within the first 20 percent of the time it would take to collect the whole amount. So it is more cost efficient for the IRS to be satisfied with the 80 percent from individual audits while using the time saved to conduct a larger number of audits. This is expected to prove less annoying to those individuals who are audited, while the larger number of audits is intended to deter tax cheating due to the increased risk of being audited (Brown 2007).

Details of Data Collection: The Evaluation Design

The details of data collection are a distinct part of the framework; they must be attended to, but they shouldn't clutter the focused outcome statement. Unfortunately, I've found that people can get caught up in the details of refining methods and lose sight of the outcome. The details typically get worked out after the other parts of the framework have been conceptualized. Details include answering the following kinds of questions:

- What existing data will be used and how will they be accessed? Who will collect new indicators data?
- Who will have oversight and management responsibility for data collection?
- How often will indicators data be collected? How often reported?
- Will data be gathered on all program participants or only a sample? If a sample, how selected?
- How will findings be reported? To whom? In what format? When? How often?

These pragmatic questions put flesh on the bones of the outcomes framework. They are not simply technical issues, however. How these questions get answered will ultimately determine the credibility and utility of the entire approach. Primary intended users need to be involved in making decisions about these issues to ensure that they feel ownership of and responsibility for all aspects of the evaluation.

How Results Will Be Used

The final element in the framework is to make sure that the data collected on the outcomes identified will be useful. One way to do this is to engage intended users in a simulation exercise in which the evaluator fabricates some potential results and intended users practice interpreting and using those results. The evaluation facilitator asks, "If the results came out this way, what would you do? If the findings came out this other way, what would that tell you, and what actions would you take? Given what you want the evaluation to accomplish, have we focused on the right outcomes and useful indicators?"

(Chapter 13 will discuss this simulation approach in greater depth.) At every stage of a utilization-focused evaluation, the evaluator facilitator pushes intended users to think seriously about the implications of design and measurement decisions for use.

Interconnections among the Distinct Parts of the Framework

The utilization-focused outcomes framework, as just reviewed, consists of six parts: a specific participant target group, a desired outcome for that group, one or more outcome indicators, a performance target (if appropriate and desired), details of data collection, and specification of how findings will be used. While these are listed in the order in which an evaluation typically facilitates the work with intended users, the conceptualization process is not linear. Groups often go back and forth in iterative fashion. The target group may not become really clear until the desired outcome is specified or an indicator designated. Sometimes formulating the details of data collection will give rise to new indicators, and those indicators force a rethinking of how the desired outcome is stated. The point is to end up with all elements specified, consistent with each other, and mutually reinforcing. That doesn't necessarily mean marching through the framework lockstep, but it does mean eventually determining all six elements.

Exhibit 7.5 provides an example of all the elements specified for a parenting program aimed at high school–age mothers.

EXHIBIT 7.5

Example of a Fully Specified Utilization-Focused Outcomes Framework

Target subgroup:	Teenage mothers at Central High School
Desired outcome:	Appropriate parenting knowledge, behaviors, and practices
Outcome indicator:	Score on Parent Practice Inventory (knowledge and behavior measures)
Data collection:	Pre- and post-test, beginning and end of program; 6-month follow-up; district evaluation office will administer and analyze results
Performance target:	75 percent of entering participants will complete the program and attain a passing score on both the knowledge and behavior scales
Use:	The evaluation advisory task force will review the results (principal, two teachers, two participating students, one agency representative, one community representative, an associate superintendent, one school board member, and the district evaluator). The task force will decide if the program should be continued at Central High School and expanded to other district high schools. A recommendation will be forwarded to the superintendent and school board.

Completing the framework often takes several tries. Exhibit 7.6 shows three versions of the utilization-focused outcomes framework as it emerged from the work of a developmental disabilities staff group. Their first effort yielded a service-oriented goal. They revised that with a focus on skill enhancement. Finally, they agreed on a meaningful client outcome: functioning independently.

EXHIBIT 7.6

Three Versions of an Outcome-Focused Framework

	Target Population: Children with Development Disabilities		
	Desired Outcome	*Outcome Indicator*	*Method*
First draft (Service-oriented)	Children with developmental disabilities will receive supportive services for improved functioning in basic daily living skills	Track hours of supportive services received, and levels and amounts of client participation in training	Case records monitoring services and participation will be aggregated quarterly
Revised (Skills-focused; interim outcome)	Children with developmental disabilities will increase their skills for functioning independently	Changes in skills on a staff assessment form	Quarterly administration of skills assessment form as part of ongoing training
Final version (Primary desired outcome)	Children with developmental disabilities will function independently in their activities of daily living	Activities of Daily Living (ADL) behavioral assessment instrument	Quarterly administration of ADL to all children in the program. Compare scores over time. Do both individual case profiles and aggregate results by categories of severity and age

A Utilization-Focused Process for Developing Outcomes

A central issue in implementing an outcomes evaluation approach is who will be involved in the process of developing the outcomes. When the purpose is ongoing management by outcomes, the program's executives and staff must buy into the process. Who else is involved is a matter of political judgment. Those involved will feel the most ownership of the resulting system.

Some processes involve only managers and directors. Other processes include advisory groups from the community. Collaboration between funders and service providers in determining outcomes is critical where contracts for services are involved. Advice from some savvy foundation funders is to match outcomes evaluation to the stage of a program's development (Daniel 1996), keep the context long term (McIntosh 1996) and "turn outcome 'sticks' into carrots" (Leonard 1996:46).

Exhibit 7.7 shows the stages of a utilization-focused approach to developing an outcomes-based management system for a program. Critical issues and parallel activities are shown for each stage. Those to be involved will need training and support. I've found it helpful to begin with an overview of the purpose of an outcomes-focused programming approach: history, trends, the political climate, and potential benefits. Then I have participants work in small groups working on the elements of the utilization-focused outcomes framework (see Exhibit 7.5) for an actual program with which they're familiar. Facilitation, encouragement, and technical assistance are needed to help such groups successfully complete the task. Where multiple groups are involved, I like to have them share their work and the issues that emerged in using the outcomes framework.

Dealing with Concerns

It's important that those involved get a chance to raise their concerns openly. There's often suspicion about political motives. Service providers worry about funding cuts and being held accountable for things they can't control. Administrators and directors of programs worry about how results will be used, what comparisons will be made, and who will control the process. Line staff worry about the amount of time involved, paperwork burdens, and the irrelevancy of it all. State civil servants responsible for reporting to the Legislature worry about how data can be aggregated at the state level. These and other concerns need to be aired and addressed. Having influential leaders visibly involved in the process enhances their own understanding and commitment while also sending signals to others about the importance being placed on outcomes.

Ten Principles for Meaningful and Useful Goals

With the utilization-focused outcomes framework as background, here are 10 principles for working with intended users to identify meaningful and useful goals.

1. *Identify outcomes that matter to those who do the work.* Outcomes and performance measures often look and feel like academic exercises to those involved. They think they're involved in a paperwork process to please some funder. But a meaningful outcomes statement articulates fundamental values. This means asking program stakeholders not only what their goals are but *why* they care about those goals (Friedman, Rothman, and Withers 2006). Goals should remind practitioners why they get up in the morning and go to work: to protect children, feed the hungry, fight disease, reduce the ravages of poverty, and house the homeless. It should be inspirational. As the great management guru Peter Drucker (2000) said when asked about the bottom line for not-for-profit organizations, "The end results are people with changed lives." Outcomes should specify how lives will be changed.

EXHIBIT 7.7

Developing a Utilization-Focused System for Managing Outcomes: Stages, Issues, and Activities

	1	2	3	4	5	6	7	8	9	10	11	12	13
Stages	Identify and engage key actors and leaders whose commitment and support will be needed for transition to management and accountability based on client outcomes	Key actors and leaders: Commit to establish a client outcome approach. Understand principles, purposes, and implications of change	Agree on intended use by intended users	Conceptualize client outcomes; select indicators; set targets	Engage line staff: Facilitate their understanding and buy-in	Design data collection system; finalize methods; pilot; establish baselines	Implement data collection; train staff and managers for data collection and use	Prepare for use: Determine management uses; potential actions; decision options and parameters, and accountability report format	Analyze results: Compare results to baseline and targets	Involve key stakeholders in processing the findings	Judge performance and effectiveness	Make management decisions. Report results	Review and evaluate the outcome-based management system
Issues	Who are the key actors and leaders who must buy in? How widespread should initial involvement be?	What level of commitment and understanding is needed? By whom? How to distinguish real commitment from mere rhetoric?	What uses are possible? What uses are doable?	What target groups? How many outcomes? What are the really important bottomline outcomes?	What are staff's perspective/ history/ concerns/ incentives?	What can be done with existing data? What new data will be needed? How can the system be integrated?	What resources will be available to support data collection? How will validity and reliability be addressed?	What incentives exist for managers to participate? How will managers be brought along? Trained? Rewarded? Who determines accountability reporting approaches?	Who will do the analysis? What additional data are needed to interpret the outcome results (e.g., demographics)	How do you keep key stakeholders engaged?	How clear are the data to support solid judgments?	What are the links between internal and external uses and audiences?	What should the system accomplish? Who determines success?
Activities	Establish leadership group	Leadership group makes strategic decision about how best to proceed and who to involve	Map out users and uses. Set priorities	Establish work team to determine outcomes; involve advisory to bring them along	Conduct staff workshops/ training	Work team to make design decisions	Collect data; pilot-test and monitor data collection	Conduct training and management team sessions based on data use and simulations and mock scenarios	Analyze data; prepare graphics	Facilitate meeting of key stakeholders	Facilitate key stakeholders in judging and interpreting	Write report; present data; facilitate management decision making	Assemble a review team of management system users and key stakeholders

2. *Distinguish between outcomes and activities.* Outcomes describe desired impacts of the program on participants: Students will read with understanding. Participants will stop smoking. Activity goals describe *how* outcome goals will be achieved: Students will read 2 hours a day. Participants will openly discuss their dependence on cigarettes. People in the program will be treated with respect.

Outcome goals should clearly state how people will be different as a result of a program. Program staff may write goals describing intended activities thinking that they have stated desired outcomes. An agricultural extension agent told me his goal was "to get 50 farmers to participate in a farm tour." But what, I asked, did he want to result from the farm tour? After some dialogue, it became clear that the desired outcome was this: "Farmers will adopt improved milking practices in their own farm operations, and thus have more income."

A corporation stated one of its goals for the year as "establishing a comprehensive energy conservation program." After we discussed that it was perfectly possible to establish such a program without ever saving any energy, they rewrote the goal: "The corporation will significantly reduce energy consumption."

3. *Specificity matters. More specific outcomes are more powerful.* Some goal statements are amazingly adept at saying nothing. I worked with a school board whose overall goal was, "Students will learn." There is no way *not* to attain this goal. It is the nature of the species that young people learn. Fortunately, they can learn in spite of the schools. The issues are *what* and *how much* they will learn from schooling.

Another favorite is "increasing awareness." It's fairly difficult to put people through 2 weeks of training on some topic (e.g., chemical dependency) and *not* increase awareness. Under these conditions, the goal of "increasing awareness of chemical dependency issues" is hardly worth aiming at. Further dialogue revealed that the program staff wanted to change knowledge, attitudes, and behaviors. Specific outcomes identify what knowledge, what attitudes, and what behaviors.

4. *Each goal should contain only one outcome.* There is a tendency in writing goal statements to overload the content. "Our program will help parents with employment, nutrition, health, and parenting needs so that their children do well in school and reach their full potential, and families are integrated into neighborhoods feeling safe and connected, and being productive." Now there's a goal written by committee with a little bit for everyone. Ten different possible outcomes are implied for three different target populations in that statement. For evaluation purposes, they must be separated.

5. *Outcome statements should be understandable.* Goals should communicate a clear sense of direction. Avoid difficult grammatical constructions and complex interdependent clauses. Goal statements should also avoid internal program or professional jargon. The general public should be able to make sense of goals. Consider these two versions of goal statements for what amount to the *same* outcome:

(a) To maximize the capabilities of professional staff and use taxpayer resources wisely while engaging in therapeutic interventions and case management processes so that children's development capacities are unencumbered by adverse environmental circumstances or experiences.

(b) Children will be safe from abuse and neglect.

Now, see if you can make sense of this beauty from the National Council of Teachers of English and the International Reading Association: "Students employ a wide range of strategies as they write and use different writing process elements appropriately to communicate with different audiences for a variety of purposes." The *New York Times* (1996) found this goal less than inspiring or user-friendly, and editorialized: "a fog of euphemism and evasion" (p. A24). Bumper sticker: Honk if you use *writing process elements* appropriately.

6. *Formal goals statements should focus on the most important program outcomes.* Writing goals should not be a marathon exercise in seeing how long a document one can produce. As human beings, our attention span is too short to focus on long lists of goals and objectives. Limit them to outcomes that matter and for which the program intends to be held accountable.

7. *State intended outcomes separately from how they are to be attained.* An agricultural extension program posited this goal: "Farmers will increase yields through the education efforts of extension, including farm tours, bulletins, and related activities." Everything after the word *yields* describes how to goal is to be attained. Keep the statement clear and crisp—focused on the intended outcome.

8. *Separate goals from indicators.* Advocates of *management by objectives* and *behavioral objectives* often place more emphasis on measurement than on establishing a clear sense of direction (e.g., Combs 1972). *Desired outcome:* All children will be immunized against polio. Indicator: Health records when children enter school show that they received 4 doses of IPV: a dose at 2 months; at 4 months; at 6 to 18 months; and a booster dose at 4 to 6 years.

9. *Thou shalt not covet thy neighbor's goals and objectives.* Goals and objectives don't travel very well. They often involve matters of nuance. It is worth taking the time for primary stakeholders to construct their own goals so that they reflect their own values, expectations, and intentions in their own language. Buy-in happens through engagement.

10. *Help all involved keep their eyes on the prize. Use outcome statements to stay focused on achieving results.* Goals clarification should be an invigorating process of prioritizing what those involved care about and hope to accomplish (see Item 1 above). Goals should not become a club for assaulting staff but a tool for helping staff focus and realize their ideals. Too often outcomes are written into proposals and reports, then forgotten. Make monitoring outcomes attainment part of staff meetings. Find out if it's true that what gets measured gets done. Orient new staff members to the program's outcome commitments. Staff should share intended outcomes with participants so that everyone knows what's expected and envisioned. Report outcomes in newsletters and other program communications. Revisit outcomes at annual retreats. An informative and revealing exercise can be to conduct an outcomes communications audit: Where in the life and work of the program are priority outcomes shared and used?

There are exceptions to all these guidelines. For example, contrary to the ninth principle, one option in working with groups is to have them review the goals of other programs, both as a way of helping stakeholders clarify their own goals and to

get ideas about format and content. From this beginning point, those in a particular situation can fine-tune others' goals to fit their values and context.

Where there is the time and inclination, then, I prefer to have key program people work on their own outcomes framework, including identifying indicators and uses of monitoring data, so that participants feel ownership and understand what commitments have been made. This can be part of the training function served by evaluators, increasing the likelihood that staff will internalize the evaluative thinking embedded in a utilization-focused outcomes framework (see Exhibit 7.5).

African Standards as an Example of Adapted Outcomes

An example of adapting goals to fit specific values and a particular context in evaluation are reviews being undertaken by national evaluation associations around the world to examine the Joint Committee Standards for Evaluation, originally formulated for educational evaluation in the United States, and adapt the standards to their own cultural contexts. The standards are, in a sense, the desired outcomes for evaluation. The utility standards, for example, "are intended to ensure that an evaluation will serve the information needs of intended users" (Joint Committee 1994:U). Here's an example of how the African Evaluation Association (AfrEA) adapted the standards to African concerns. One of the U.S. accuracy standards reads,

> **Justified Conclusions** The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them. (A10)

The African Evaluation Association adopted revised standards at its conference in Niamey, Niger, in January 2007. They revised the "Justified Conclusions" standard into two parts:

> A10a. **Relevant Conclusions.** The conclusions of an evaluation should result from methods and analysis so that stakeholders can appreciate them in full objectivity.
>
> A10b. **Realistic recommendations reached by consensus.** The recommendations of an evaluation should be validated by stakeholders, feasible and linked to expected results. (AfrEA 2007:A10a, A10b)

The original Joint Committee Standards didn't address recommendations. The African evaluators sought input from evaluators and national evaluation associations throughout Africa and determined that they needed a standard regarding recommendations. In this way, they made the standards their own while also affirming and adopting the rest of the Joint Committee framework and standards. They also wrote an important contextual introduction to the standards to delineate their relevance to and importance for Africa.

Performance Indicators

As we have seen in this chapter, discussions of goals quickly turn to how to measure attainment of goals and that leads to performance indicators. Performance indicators have become so important and widely used that they deserve some additional consideration, including how they relate to broader evaluation concerns. Chapter 4 discussed monitoring as one use of data in monitoring and evaluation (M&E) systems. Performance indicators have become central to such approaches. Ongoing monitoring of indicators against desired target levels may be called performance

measurement or performance monitoring. Because of the increased importance of performance monitoring in the United Kingdom, the British Commonwealth, and the European Union, the Royal Statistical Society (2003) created a special group to review the quality of the United Kingdom's system. In a report titled *Performance Indicators: Good, Bad, and Ugly,* the reviewers began by examining the multiple purposes of performance monitoring: (1) to assess the impact of government policies on services, (2) to identify well-performing or underperforming institutions and public servants, and (3) for public accountability. Hence, government is in the position of both monitoring public services and being monitored, itself, by performance indicators. This makes the political stakes quite high. "Performance monitoring done well is broadly productive for those concerned. Done badly, it can be very costly and not merely ineffective but harmful and indeed destructive" (Royal Statistical Society 2003:2).

The potential positive contribution of performance monitoring is captured in the mantra that *what gets measured gets done.* Well-developed and appropriate indicators both focus attention on priority outcomes and provide accountability for achieving those outcomes. The shadow side of performance indicators is that measuring the wrong thing means the wrong thing gets done. Consider the goal of protecting children in child protection agencies. Children who are neglected or abused are removed and placed in foster homes. One common goal of such agencies, in an effort to support families and as a matter of public policy, is to reunite children with their natural families. In many states, this is monitored through a reunification rate. When comparative data are made public for accountability purposes, counties or agencies with low reunification rates experience pressure to raise their rates. That's where the problem arises. It's an easy rate to manipulate. More children can simply be reunited with their natural families. But the real goal is to *safely* reunite children with their families. That involves a different indicator.

The Royal Statistical Society report recommends that before introducing performance monitoring in any public service, a performance monitoring (PM) protocol should be established. This is an orderly record not only of decisions made but also of the reasoning or calculations that led to those decisions. A PM protocol should cover objectives, design considerations, and the definition of performance indicators, sampling versus complete enumeration, the information to be collected about context, the likely perverse behaviors or side effects that might be induced as a reaction to the monitoring process, and also the practicalities of implementation. Procedures for data collection, analysis, presentation of uncertainty and adjustment for context, together with dissemination rules, should be explicitly defined and reflect good statistical practice. Because of their usually tentative nature, performance indicators should be seen as "screening devices" and not overinterpreted. If quantitative performance targets are to be set, they need to have a sound basis, take account of prior (and emerging) knowledge about key sources of variation, and be integral to the PM design (p. 2).

The report emphasizes the importance of education and training in the appropriate use and interpretation of performance indicators. Special emphasis is given to the importance of *independent scrutiny* as a safeguard of public accountability, methodological rigor, and to be fair to individuals and/or institutions being monitored. "The scrutineers' role includes

checking that the objectives of Performance Monitoring are being achieved without disproportionate burden, inducement of counterproductive behaviours, inappropriate setting or revision of targets, or interference in, or over-interpretation of, analyses and reporting" (p. 4).

Concerns about the misuse of performance indicators follow from Campbell's Law, formulated by Donald Campbell, one of evaluation's most distinguished pioneers: "The more any quantitative social indicator is used for social decision-making," he posited, "the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell 1988:360). Consider this example:

> Police officers in New Orleans manipulated crimes statistics to make it look like the crime rate was going down stimulated by the Department's policy of handing out awards to leaders of districts with the lowest crime statistics. Five police officers were fired over the scandal. (Associated Press, October 23, 2003)

Sometimes the "corruption" (or at least distortion) begins in the definition and labeling of indicators. In the U.S. Senate debate over whether to allow drilling for oil in the Alaskan wilderness, the issue arose about what it means for an area to be a wilderness. The traditional definition has included that a wilderness is "roadless." The Bush Administration's U.S. Department of Interior revised the definition of roadless as follows: "The term 'roadless' does not mean an absence of roads. Rather, it indicates an attempt to minimize the construction of permanent roads." Thus, gravel roads constructed for drilling and logging don't count and areas with such roads remain counted in the statistics on how much of the wilderness is "roadless" (Barringer 2005:16).

Such examples make "the performance movement" controversial and in need of ongoing evaluation itself (Radin 2006).

Campbell's Law

"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

—Donald Campbell (1988:360)

Performance Indicators and Evaluation

Performance monitoring based on *key performance indicators* (KPIs) has become part of the political culture. Governments and politicians are expected to set targets and report on progress as a basis for public accountability. The usefulness of performance indicators depends on their credibility, relevance, validity, transparency, and meaningfulness—and an appropriate and fair process for interpreting them. The Royal Statistical Society (2003) report provides excellent guidance in this regard. What it does not do is distinguish monitoring from evaluation.

Performance indicators are one tool in a very large evaluation toolkit that includes a wide variety of methods, data collection techniques, measures, and models. Given the rapid and widespread proliferation of performance-monitoring approaches, there is the danger that many will think that performance measurement is sufficient for or equivalent to evaluation. But performance measurement merely portrays trends and directions. Indicators tell us

whether something is increasing, declining, or staying the same. Evaluation takes us deeper into asking why indicators are moving in the direction they are, how the movement of indicators are related to specific interventions, what is driving the movement of indicators, and what values should guide interpretation of indicators in making judgments. Utilization-focused performance measurement adds the importance of being clear about the primary intended users and intended uses of performance indicators.

Burt Perrin has been one of evaluation's most thoughtful theorists about the relationship between performance monitoring and evaluation. Exhibit 7.8 presents his principles and strategies for effective use of performance measures. Given the increasing importance of performance measurement in the public sector around the world, evaluators, policymakers, and the general public need to understand both the strengths and limitations of key performance indicators.

EXHIBIT 7.8

Strategies for Effective Use of Performance Measures

- Recognize that performance indicators are most appropriate for use in planning and monitoring, not for evaluation.
- Performance measures can serve as a means of identifying where more comprehensive evaluation approaches would be most useful, in this way, help in responsible allocation of resources for evaluation.
- Recognize that every evaluation method, including performance indicators, has limitations that can only be overcome by using a combination of methods . . . . It is usually necessary to balance quantitative data from performance indicators with qualitative forms of information for balance.
- Be strategic, recognizing that performance indicators are appropriate for some activities and not for others.
- Be realistic about the context, such as political or organizational requirements for performance indicators, irrespective of their appropriateness, and help programs work around these realities.
- Make sure the measurements are at the right level. The logic model can help identify what forms of outcomes may be realistically expected given the status of the program. Do not hold programs accountable for impacts that are unrealistic.
- Always test indicators in advance.
- Anticipate misapplications.
- Review, revise, and update measures frequently. Don't expect that new measures will be perfect the first time around.
- Actively involve stakeholders, including program staff and consumers, in developing, reviewing, and revising measures and actively involve them in interpreting findings and identifying implications.
- Use multiple indicators, in order to examine a variety of program aspects, including process as well as outputs and outcomes.

—Burt Perrin (1998:375–76)

Levels of Goal Specification

From Overall Mission to Specific Objectives

To facilitate framing evaluation questions, evaluators may have to work with primary stakeholders to clarify purposes at three levels: (1) the overall mission of the program or organization, (2) the goals of specific programmatic units (or subsystems), and (3) the specific objectives that specify desired outcomes. The mission statement describes the general direction of the overall program or organization in long-range terms. The peacetime mission of the U.S. Army is simply "Readiness." A mission statement may specify a target population and a basic problem to be attacked. For example, the mission of the Minnesota Comprehensive Epilepsy Program was to "improve the lives of people with epilepsy."

The terms *goals* and *objectives* have been used interchangeably up to this point, but it is useful to distinguish between them as representing different levels of generality. Goals are more general than objectives and encompass the purposes and aims of program subsystems (i.e., research, education, and treatment in the epilepsy example). Objectives are narrow and specific, stating what will be different as a result of program activities, that is, the concrete outcomes of a program. (Note: In some parts of the world, objectives are more general and goals are more specific.) To illustrate these differences, a simplified version of the mission statement, goals, and objectives for the Minnesota Comprehensive Epilepsy Program is presented in Exhibit 7.9. This outline was developed after an initial discussion with the program director. The purpose of the outline was to establish a context for later discussions aimed at more clearly framing specific evaluation questions. In other words, we used this goals clarification and objectives *mapping exercise* as a means of focusing the evaluation question rather than as an end in itself.

The outline of goals and objectives for the Epilepsy Project (Exhibit 7.9) illustrates several points. First, the only dimension that consistently differentiates goals and objectives is the relative degree of specificity of each: Objectives narrow the focus of goals. There is no absolute criterion for distinguishing goals from objectives; the distinction is always a relative one.

Second, this outline had a specific evaluation purpose: to facilitate priority setting as I worked with primary intended users to focus the evaluation. Resources were insufficient to fully evaluate all three component parts of the program. Moreover, different program components faced different contingencies. Treatment and research had more concrete outcomes than education. The differences in the specificity of the objectives for the three components reflected real differences in the degree to which the content and functions of those program subsystems were known at the beginning of the evaluation. Thus, with limited resources and variations in goal specificity, it was necessary to decide which aspects of the program could best be served by evaluation.

Third, the outline of goals and objectives for the Comprehensive Epilepsy Program is not particularly well written. I constructed the outline from notes taken during my first meeting with the director. At this early point in the process, the outline was a tool for posing this question to evaluation decision makers: *Which program components, goals, and objectives should be evaluated to produce the most useful information for program improvement and decision making?* That is the question. To answer it, one does not need technically perfect goal

EXHIBIT 7.9

Minnesota Comprehensive Epilepsy Program: Mission Statement, Goals, and Objectives

Program Mission: Improve the lives of people with epilepsy

Research Component

Goal 1: Produce high-quality, *scholarly research* on epilepsy

Objective 1: Publish research findings in high-quality, refereed journals

Objective 2: Contribute to knowledge about

a. neurological aspects of epilepsy
b. pharmacological aspects of epilepsy
c. epidemiology of epilepsy
d. social and psychological aspects of epilepsy

Goal 2: Produce interdisciplinary research

Objective 1: Conduct research projects that *integrate* principal investigators from different disciplines

Objective 2: Increase meaningful *exchanges* among researchers from different disciplines

Education Component

Goal 3: Health professionals will know the nature and effects of epilepsy behaviors

Objective 1: Increase the *knowledge* of health professionals who serve people with epilepsy so that they know

a. what to do if a person has a seizure
b. the incidence and prevalence of epilepsy

Objective 2: Change the attitudes of health professionals so that they

a. are sympathetic to the needs of people with epilepsy
b. believe in the importance of identifying the special needs of people with epilepsy

Goal 4: Educate persons with epilepsy about their disorder

Goal 5: Inform the general public about the nature and incidence of epilepsy

Treatment Component

Goal 6: Diagnose, treat, and rehabilitate persons with severe, chronic, and disabling seizures

Objective 1: Increase seizure control in treated patients

Objective 2: Increase the functioning of patients

statements. Once the evaluation is focused, relevant goals and objectives can be reworked as necessary. The point is to avoid wasting time in the construction of grandiose, complicated models of program goals and objectives just because the folklore of evaluation prescribes such an exercise. In comprehensive, multidimensional programs, evaluators can spend so much time working on goals statements that considerable momentum is lost.

Establishing Priorities: Importance versus Utility

Let me elaborate the distinction between writing goals for the sake of writing goals and writing them to use as tools in narrowing the focus of an evaluation. In utilization-focused evaluation, goals are prioritized in a manner quite different from that usually prescribed. The classic criterion for prioritizing goals is ranking or rating in terms of *importance* (Edwards, Guttentag, and Snapper 1975; Gardiner and Edwards 1975). The reason seems commonsensical: Evaluations ought to focus on important goals. But from a utilization-focused perspective, what appears to be most sensible may not be most useful.

The most important goal may not be the one that decision makers and intended users most need information about. In utilization-focused evaluation, goals are also prioritized on the basis of what information is most needed and likely to be most useful, given the evaluation's purpose. For example, a final end-of-project summative evaluation would likely evaluate goals in order of overall importance, but a formative (improvement-oriented) evaluation might focus on a goal of secondary importance because it is an area being neglected or proving particularly troublesome.

Ranking goals by importance is often quite different from ranking them by the utility of evaluative information needed at a particular time. Exhibit 7.10 provides an example from the Minnesota Comprehensive Epilepsy Program, contrasting goals ranked by importance and utility. Why the discrepancy? The staff did not feel they needed a formal, external evaluation to monitor attainment of the most important program goal. The publishing of scholarly research in refereed journals was so important that the director was committed to personally monitor performance in that area. Moreover, he was relatively certain about how to achieve and measure that outcome, and he had no specific evaluation question related to that goal that he needed answered. In contrast, the issue of comprehensiveness was quite difficult to assess. It was not at all clear how comprehensiveness could be facilitated, although it was third on the importance list. Data on comprehensiveness had high formative utility, and this became the priority focus for the formative evaluation.

The education goal, second on the usefulness list, does not even appear among the top four goals on the importance list. Yet information about educational impact was ranked high on the usefulness list because it was a goal area about which the program staff had many questions. The education component was expected to be a difficult, long-term effort. Information about how to increase the educational impact of the Comprehensive Epilepsy Program had high use potential. In a utilization-focused approach, the primary intended users make the final decision about evaluation priorities.

In my experience, the most frequent reason for differences in importance and usefulness rankings is variation in the degree to which decision makers already have what

EXHIBIT 7.10

Minnesota Comprehensive Epilepsy Program: Goals Ranked by Importance to Program versus Goals Ranked by Utility of Evaluative Information Needed by Primary Users

Ranking of Goals by Program Importance	*Ranking Goals by Need for and Usefulness of Evaluative Information to Primary Intended Users*
1. Produce high-quality scholarly research on epilepsy	1. Integrate the separate program components into a comprehensive whole that is greater than the sum of its parts
2. Produce interdisciplinary research	2. Educate health professionals about epilepsy
3. Integrate the separate components into a whole	3. Diagnose, treat, and rehabilitate people with chronic and disabling seizures
4. Diagnose, treat, and rehabilitate people with chronic and disabling seizures	4. Produce interdisciplinary research

they consider good information about performance on the most important goal and the overall purpose of the evaluation (formative vs. summative vs. developmental vs. knowledge-generating). At the program level, staff members may be so involved in trying to achieve their most important goal that they are relatively well informed about performance on that goal. Performance on less important goals may involve less certainty for staff; information about performance in that goal area is therefore more useful for improvement because it tells staff members something they do not already know. On the other hand, for summative evaluations aimed at funders, they will typically want to know about attainment of the most important goals.

What I hope is emerging through these examples is an image of the evaluator as an active-reactive-adaptive problem solver. The evaluator actively solicits information about program contingencies, organizational dynamics, environmental uncertainties, and decision makers' goals in order to focus the evaluation on questions of real interest and utility to primary intended users at a particular stage in the life of the program and for a specific evaluation purpose.

Evaluation of Central versus Peripheral Goals

Prioritizing goals on the basis of perceived evaluative utility means that an evaluation might focus on goals of apparent

peripheral importance rather than more central program goals. This is a matter of some controversy. In her early work, Weiss (1972b) offered the following advice to evaluators:

> The evaluator will have to press to find out priorities—which goals the staff sees as critical to its mission and which are subsidiary. But since the evaluator is not a mere technician for the translation of a program's stated aims into measurement instruments, he has a responsibility to express his own interpretation of the relative importance of goals. *He doesn't want to do an elaborate study on the attainment of minor and innocuous goals* [italics added], while some vital goals go unexplored. (Pp. 30–31)

Contrast that advice with the perspective of an evaluator from our study of use of federal health evaluations:

> I'd make this point about minor evaluation studies. If you have an energetic, conscientious program manager, he's always interested in improving his program around the periphery, because that's where he usually can. And an evaluation study of some minor aspect of his program may enable him to significantly improve. [EV52:171]

In our study, we put the issue to decision makers and evaluators as follows:

> Another factor sometimes believed to affect use has to do with whether the central objectives of a program are evaluated. Some writers argue that evaluations can have the greatest impact if they focus on major program objectives. What happened in your case?

The overwhelming consensus was that, at the very least, central goals ought to be evaluated and, where possible, both central and peripheral goals should be studied. As they elaborated, nine decision makers and eight evaluators said that utilization had probably been increased by concentrating on *central issues.* This phrase reflects an important shift in emphasis. As they elaborated their answers about evaluating central versus peripheral goals, they switched from talking about goals to talking about "issues." Utilization is increased by focusing on central issues. *And what is a central issue? It is an evaluation question that someone really cares about.* The subtle distinction here is critical. Evaluations are useful to decision makers if they focus on central *issues*—which may or may not include evaluating attainment of central goals.

The Personal Factor Revisited

Different people will have different perceptions of what constitutes central program goals or issues. Whether it is the evaluator's opinion about centrality, the funder's, some special interest group's perspective, or the viewpoints of program staff and participants, the question of what constitutes central program goals and objectives remains an intrinsically subjective one. It cannot be otherwise. The question of central versus peripheral goals cannot really be answered in the abstract. The question thus becomes central from whose point of view? The personal factor (Chapter 3) intersects the goals clarification process in a utilization-focused evaluation. Increasing use is largely a matter of matching: getting information about the right questions, issues, and goals to the right people.

Earlier in this chapter, I compared the goals clarification process with the party game of Twenty Questions. Research indicates that different individuals behave quite differently in such a game (and, by extension, in any decision-making process).

In a classic experiment, Worley (1960) studied subjects' information-seeking endurance in the game under laboratory conditions. Initially, each subject was presented with a single clue and given the option of guessing what object the experimenter had in mind or of asking for another clue. This option was available after each new clue, but a wrong guess would end the game. *Worley found large and consistent individual differences in the amount of information players sought.* Such research provides evidence that decision-making and problem-solving behavior is dynamic, highly variable, and contingent on both situational and individual characteristics. This does not make the evaluator's job any easier. It does mean that the personal factor remains the key to evaluation use. The careful selection of knowledgeable, committed, and information-valuing people makes the difference. The goals clarification game is most meaningful when played by people who are searching for information because it helps them focus on central issues without letting the game become an end in itself or turning it into a contest between staff and evaluators.

Global Political Goals

The languages of goals and performance indicators have become part of the global political culture. The Kyoto Climate Treaty set specific targets for reduction of greenhouse gases. Most industrialized nations are required to cut emissions below 1990 levels (although some will be allowed to increase emissions by up to 10 percent over 1990 levels). In general, developing nations have no obligation to cut emissions now, but may be asked to make future cuts. Overall, the protocol's goal is to reduce carbon emissions by 5 percent below 1990 levels by 2008–2012, with further reductions to be negotiated in the future. No one expects these goals to be met. Their purpose and use is to focus the international political dialogue. The goals are important symbols that call attention to global warming and provide a basis for monitoring as the political dialogue continues.

In July 2007, international organizations, public officials, statistical agencies, academics, leaders of civil society, business representatives, and media gathered in Istanbul for a World Forum on Measuring and Fostering the Progress of Societies sponsored by the Organisation for Economic Cooperation and Development. The purpose of the conference was to focus worldwide attention on how societies can better use indicators to support dialogue and debate about what "progress" means. Participants affirmed the importance of measuring and fostering the progress of societies in all their dimensions, not just economic, but also social well-being, health status, and environmental quality. The conference called for producing high-quality, evidence-based information that can be used to form a shared view of societal well-being and its improvement over time.

Likewise the UN Millennium Development Goals serve the purpose of focusing international development efforts and providing a shared platform and language for political dialogue. Substantial resources have been committed to data collection for monitoring progress toward the goals. The World Bank and other international agencies have built strategic plans around the goals. Yet no one can realistically expect these goals to be achieved—not even close.

These are examples of the *process use* of goals on the global stage. They are used to provide a framework for communications and bring focus to widely disparate development efforts. They reinforce the

The UN Millennium Development Goals

Goal 1: Eradicate extreme poverty and hunger
- Reduce by half the proportion of people living on less than a dollar a day
- Reduce by half the proportion of people who suffer from hunger

Goal 2: Achieve universal primary education
- Ensure that all boys and girls complete a full course of primary education

Goal 3: Promote gender equality and empower women
- Eliminate gender disparity in primary and secondary education, preferably by 2005, and at all levels by 2015

Goal 4: Reduce Child mortality
- Reduce by two thirds the mortality rate among children under five

Goal 5: Improve maternal health
- Reduce by three quarters the maternal mortality ratio

Goal 6: Combat HIV/AIDS, malaria, and other diseases
- Halt and begin to reverse the spread of HIV/AIDS
- Halt and begin to reverse the incidence of malaria and other major diseases

Goal 7: Ensure environmental sustainability
- Integrate the principles of sustainable development into country policies and programmes; reverse loss of environmental resources
- Reduce by half the proportion of people without sustainable access to safe drinking water
- Achieve significant improvement in lives of at least 100 million slum dwellers, by 2020

Goal 8: Develop a global partnership for development
- Develop further an open trading and financial system that is rule based, predictable, and non-discriminatory, includes a commitment to good governance, development, and poverty reduction—nationally and internationally
- Address the least developed countries' special needs. These include tariff-and quota-free access for their exports; enhanced debt relief for heavily indebted poor countries; cancellation of official bilateral debt; and more generous official development assistance for countries committed to poverty reduction
- Address the special needs of landlocked and small island developing States
- Deal comprehensively with developing countries' debt problems through national and international measures to make debt sustainable in the long term
- In cooperation with the developing countries, develop decent and productive work for youth
- In cooperation with pharmaceutical companies, provide access to affordable essential dugs in developing countries
- In cooperation with the private sector, make available the benefits of new technologies—especially information and communications technologies

idea that we are all part of a global community with a common vision, shared goals, and mutual accountability. They illustrate the way in which evaluative thinking has become embedded in the global political culture.

Legislating BHAGs (Big Hairy Audacious Goals)

In 2006, the California legislature passed a bill that included the goal of *eliminating all childhood poverty in California in 20 years* (Assembly Bill 2556). It was ultimately vetoed by the governor. Such grandiose goals are controversial.

Professor Deborah Kerr, a senior lecturer at the Bush School of Government and Public Service, Texas A&M University, commented on this approach to goal-setting for *Governing Magazine*. Based on years having studied goals and the ways in which they do and don't work, she said,

> If you do goal-setting right, you'll get improvements, but you need to have a goal that's specific, measurable, achievable and there has to be a time element to it. I call goals like this—and you see them throughout government—"snap goals," because people snap under the pressure of pursuing something they can't accomplish.

Governing Magazine commentators Katherine Barrett and Richard Greene added,

> These are different from so-called "stretch goals," which encourage people to work their hardest to achieve something they have some hope of actually accomplishing. Human Resource experts believe that people like being challenged, as long as there's a chance of success. But when you stretch people until they're ready to snap, there's a good chance they'll either ignore the goal altogether or grow increasingly frustrated as they face failure on a daily basis. (Barrett and Greene 2006)

However, it is important to distinguish the effects of large-scale community and national goals, such as the proposed California commitment, from goals at the program and project level, which is what Kerr, Barrett, and Greene seem to be aiming at. These are different units of analysis and serve different purposes. Program and project outcomes focus on a more direct and immediate accountability with some hoped-for direct connection between the intervention and results. National and global goals involve the accumulation and aggregate effects of numerous separate but ultimately interdependent projects, programs, initiatives, and policies. Cumulative and aggregate societal outcomes are what former U.K. Prime Minister Tony Blair had in mind when he committed to end child poverty by 2020 (Minoff 2005).

The Goals Paradox

This chapter began with an evaluation of Nasrudin's hunting trip in search of bears. For Nasrudin, that trip ended with the "marvelous" outcome of seeing no bears. Our hunting trip in search of the role of goals in evaluation has no conclusive ending because the information needs of primary intended users will vary from evaluation to evaluation and situation to situation. Focusing an evaluation on program goals and objectives is clearly not the straightforward, logical exercise depicted by the classical evaluation literature because decision making in the real world is not purely rational and logical. This is the paradox of goals. They are rational abstractions in nonrational systems. Statements of goals emerge at the interface between the ideals of human rationality and the reality of diverse human values and ways of thinking. Therein lies their strength and their weakness. Goals provide direction for action and evaluation, but only for those who share in the values expressed by the goals.

Evaluators live inside that paradox. One way out of the paradox is to focus the evaluation without making goal attainment the central issue. The next chapter considers alternatives to goals-based evaluation.

Follow-Up Exercises

1. Locate the goals and objectives of a program in your area. Review them against the criteria in this chapter. Has the program identified indicators for monitoring goal attainment? If so, assess the quality and appropriateness of the indicators. If indicators have not been identified, develop your own examples of appropriate indicators for this program's goals and objectives.

2. This exercise involves analyzing goals conflict among different stakeholders. The example that follows is aimed at helping you find and discuss *your own example* of goals conflict.

In 1980, Candy Lightner's 12-year-old daughter, Cari, was killed by a drunk driver—a repeat offender. Brought to trial, the driver was given a slap on the wrist and released. Outraged, Lightner founded Mothers Against Drunk Drivers (MADD). During the 8 years she headed MADD, she built the organization from a one-woman crusade into a worldwide movement. The goal of MADD was to reduce drunk driving traffic fatalities, and the organization has been highly effective in raising public disapproval of drunk driving. The proportion of traffic fatalities that are alcohol-related has dropped 40 percent over the past quarter century. Most observers give substantial credit for that decline to the efforts of MADD. Today MADD's effect is felt with chapters in all 50 states, all Canadian provinces, and many international affiliates. Its goals are to educate, prevent, deter, and punish. MADD helps victims, monitors the courts, and works to pass stronger antidrunk driving legislation.

But Candy Lightner has become alienated from MADD. She left in a highly visible and widely publicized display of anger and disgust from the organization that she herself created and served as founding president. Officially, she left because MADD changed its goals. "It has become far more neo-prohibitionist than I ever wanted or envisioned," she has explained. "I didn't start MADD to deal with alcohol. I started MADD to deal with the issue of drunk driving" (Westley, Zimmerman, and Patton 2006:195). She believed that if MADD really wanted to save lives, it would focus on going after the most chronic problem drunk drivers. Instead, some charge, MADD has become a prohibitionist organization, trying to completely outlaw driving after even one drink—a zero tolerance approach for everyone. Those who oppose this prohibitionist focus, including Candy Lightner, argue that it is misguided and ultimately ineffective. MADD is no longer a safety-promotion organization, they complain, but an antialcohol organization. Thus do committed stakeholders in an organization sometimes do battle with supporters and opponents about what their mission and goals are.

With this example in mind, find your own example of an issue on which major stakeholders are in conflict about priority goals and objectives. Identify and discuss the competing stakeholder positions. Discuss the evaluation implications of these different positions.

3. Find an example of a prominently publicized community indicator (such as crime statistics, test scores in school, immunization rates, employment rates, poverty rates, HIV/AIDS infection rates,

etc.) Find news media examples where changes in these rates are linked in some way to accountability. Analyze the use of such indicators in public policy discussions and political debates. Assess the appropriateness of the indicators you have identified for evaluation purposes.

4. There are eight UN Millennium Development Goals (see section Global Political Goals). A number of organizations are monitoring progress on these goals including regular reports from the United Nations. Select one of the goals and find, through a Web search, the latest indicators data on progress toward the goal you have selected. Analyze the commentary you find from experts about this goal. What is your assessment of the role such goals play in the international arena. Present and discuss the pros and cons of such global goal setting.