2

Evaluation Policy and Evaluation Practice

William M. K. Trochim

Abstract

The author develops the basic idea of evaluation policy, describes a practical model for development and revision of evaluation policies (including a taxonomy, structure, and set of principles), and suggests critical challenges and opportunities for the future of evaluation policy. An evaluation policy is any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation. Every entity that engages in evaluation, including government agencies, private businesses, and nonprofit organizations, has evaluation policies. Sometimes they are explicit and written; more often they are implicit and ad hoc principles or norms that have simply evolved over time.

WILEY InterScience®
DISCOVER SOMETHING GREAT

Evaluation policies profoundly affect the day-to-day work of all evaluators and ultimately the quality of the programs they evaluate. Many recent and current controversies or conflicts in the field of evaluation can be viewed, at least in part, as a struggle around evaluation policy. Because evaluation policies typically apply across multiple evaluations, influencing policies directly may have systemic and far-reaching effects for practice. Evaluation practice must play a critical role in informing and shaping the development of evaluation policies. © Wiley Periodicals, Inc.

Evaluation needs a fresh look at the idea of evaluation policy and its intimate, dynamic connection with evaluation practice. On the one hand, evaluation policy has always been a topic of concern in the field. One can read almost any classic evaluation text or article and cull from it some relevant implications for evaluation policies, ranging from choice of method to involvement of relevant stakeholders to the importance of addressing ethical and regulatory issues, and so on. On the other hand, most of the potential implications for evaluation policies are necessarily couched in the conditional and qualifying language of theory. One searches the evaluation literature in vain for context-specific sets of candidate policies, for guidelines on expressing such policies, and for ways of managing and evaluating them to achieve better policies over time.

The underlying and self-consciously provocative thesis of this chapter is that developing well-informed evaluation policies that can guide evaluation practice may be the most important issue currently facing our field. It's more important than getting our methods right, or concerns about validity. It's more important than the ethics of evaluation. It's more important than making evaluation participatory, or using it for empowerment. Why? Because evaluation policy encompasses all of those things and more. It touches virtually everything we think about or do in evaluation.

This chapter aims to accomplish several things. First is the obvious need to attempt a definition of evaluation policy, describe its relationship with practice, and in general identify some of the major conceptual issues involved. Second, a methodology will be offered that can be used for managing evaluation policies in contexts ranging from large and complex systems such as the federal government to local evaluations conducted in small organizations or groups. This methodological model includes a taxonomy of evaluation policy types, a structure for managing policy complexity, and a set of principles that can guide policy development. Some fragmentary and suggestive initial examples of how evaluation policies operate and are organized will be offered. Finally, some of the challenges that we face in evolving the idea of evaluation policy will be discussed and potential opportunities considered. In the end, all roads in this chapter lead back to the primary thesis: *evaluators need to get serious about evaluation policy and its relationship to practice.*

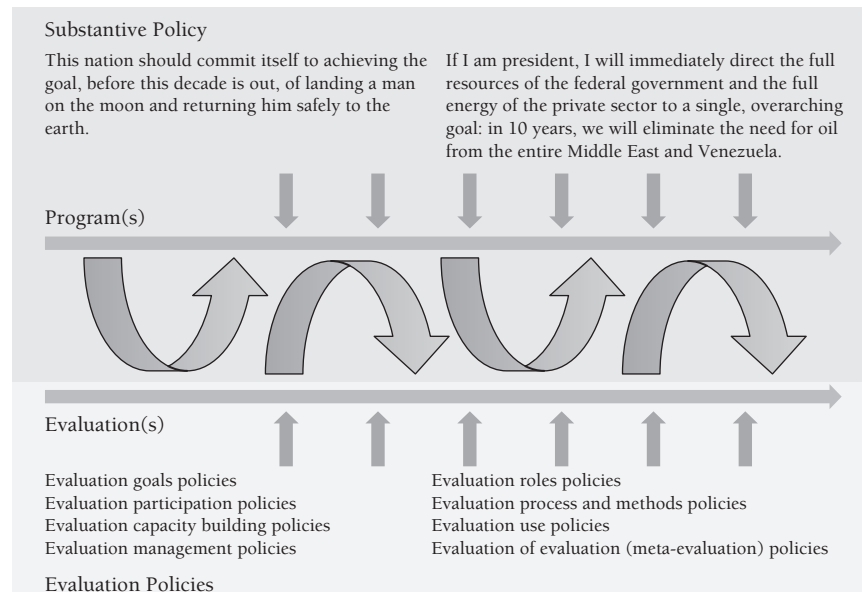## Evaluation Policy: Definitional and Conceptual Issues

A major housekeeping issue we have to address before we can get into the heart of this discussion is the confusion between evaluation policy and the traditional idea of policy. When most people think about the term *policy* they probably think first about big policy, major national policy, well-known policies that are familiar to us all. For instance, they might think about John Kennedy in 1961 saying, "This nation should commit itself to achieving the goal, before this decade is out, of landing a man on the moon and returning him safely to earth." This certainly qualifies as a major policy statement, one that successfully shaped the efforts of a generation. The statement had some of the key characteristics one might expect in any high-level general policy. It had a definable outcome in mind. It had a clear time-frame. As is typical of policies, it is a general statement that does not describe how it will be achieved operationally; Kennedy didn't talk about payloads, lunar landers, or astronauts drinking Tang. In fact, they probably didn't even have Tang at that point. Kennedy's statement can be considered a broad or high-level *substantive* policy.

Here's another example, more recent, from then-candidate Barack Obama, who in August 2008 said, "If I am president, I will immediately direct the full resources of the federal government and the full energy of the private sector to a single, overarching goal: in ten years, we will eliminate the need for oil from the entire Middle East and Venezuela." Now, that was not at the time a policy because he was not yet president, but should he carry the statement into his administration it will become a policy, a very high-level substantive policy. Notice that he doesn't talk about the details of how he is going to implement that policy. The policy simply circumscribes the direction that more detailed policy and practice would take.

So let's begin with a distinction between what will be termed here "substantive policy" and "evaluation policy," as depicted in Figure 2.1. It is tempting to use the term *public policy* rather than substantive policy, because the former is more familiar in evaluation circles, but there are policies other than just "public" ones, and the intent here is to distinguish evaluation policy from *any* kind of substantive policy, public or otherwise. When most people think of substantive policy they typically have in mind the types of statements articulated by Kennedy and Obama. Usually policies of that sort get translated into operational objects and practices such as lunar landers, Tang for astronauts, special new technologies for fuel efficiency, training and simulation programs, and so on. That is, we might call the stuff that policies are translated into its programs or practices. There is an often complex translation that occurs in operationalizing substantive policies into practices: high-level policies are typically translated into more specific subpolicies on the way to their manifestation in practice.

There is a dynamic throughout the life of a policy, and the programs or activities associated with it. Evaluation plays a critical role in providing

**Figure 2.1. The Relationship Between Substantive
and Evaluation Policy**

Substantive Policy

This nation should commit itself to achieving the goal, before this decade is out, of landing a man on the moon and returning him safely to the earth.

If I am president, I will immediately direct the full resources of the federal government and the full energy of the private sector to a single, overarching goal: in 10 years, we will eliminate the need for oil from the entire Middle East and Venezuela.

Program(s)

Evaluation(s)

Evaluation goals policies
Evaluation participation policies
Evaluation capacity building policies
Evaluation management policies

Evaluation roles policies
Evaluation process and methods policies
Evaluation use policies
Evaluation of evaluation (meta-evaluation) policies

Evaluation Policies

essential feedback about what's happening in the programs or practices or technologies associated with policies. Evaluation policies guide how evaluation happens, as depicted at the bottom of Figure 2.1 (which also foreshadows the categories of policies that we will include in our taxonomy). That is, evaluation policies can shape evaluation practice. In turn, evaluation serves programs and activities and, through them, the substantive policies that guide them. But there is a clear distinction between evaluation policies and the substantive policies they serve. With this important distinction in place, the remainder of this chapter sets aside the issue of broad substantive policy and concentrates on the topic of evaluation policy and its relationship to evaluation practice.

Definition of Evaluation Policy. What is evaluation policy? Here's a potential simple formal definition: *An evaluation policy is any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation.*

Every group and organization that engages in evaluation—including government agencies, private businesses, and nonprofit organizations—has evaluation policies. Sometimes they are written and explicit; at other times they are unwritten and implicit, ad hoc principles or norms that have simply evolved over time. The contention here is that all evaluations are *already* guided by policy, whether the protagonists involved in them recognize or acknowledge it or not. The problem is that most of the policies that guide

evaluations are unwritten. You would have a hard time pointing to them; they are not transparent either to those within the organization or to those outside. In the absence of written evaluation policies, organizations often seem to make up policies as they go along—too often without consulting others on what they have done or informing themselves about evaluation best practices.

One of the key definitional questions about policy has to do with how we distinguish it from other things that are potentially confused with it. For instance, how do we tell the difference between a policy and a guideline, a policy and a standard, or a policy and a theory? The argument here is that standards, guidelines, and theories become policies only if and when they are consciously adopted to guide decisions or actions about evaluation and when the organization institutes consequences for encouraging or enforcing them. If we have an evaluation theory or approach, such as utilization-focused evaluation (Patton, 2008), empowerment evaluation (Fetterman, Kaftarian, & Wandersman, 1996), theory-driven evaluation (Chen & Rossi, 1990), or any other of the worthy forms of evaluation, those are simply what they are: theories or approaches. They become evaluation policy when an organization decides it is going to use this approach or adopt this theory in doing evaluations; failing to do so will have consequences.

These two characteristics of policy—consciousness and consequence—are not always consistent within an organization. For instance, it may be the case that some people, such as upper management, hold evaluation policy expectations and attempt to incentivize or enforce their use while at the same time others in their organization have their own expectations, or are unaware of their views, or are uninfluenced by their efforts at enforcing them. This is one of the major justifications for encouraging that evaluation policies be written and communicated.

Why Is Evaluation Policy Important? Why should we care about written evaluation policies and how they are developed? First, evaluation policy is valuable for its signaling role. An evaluation policy can be thought of as a type of communication mechanism. It constitutes a signal to the entire organization and its stakeholders, communicating what evaluations should be done, what resources expended, who is responsible, how they should be accomplished, and so on. It can be an efficient way to communicate and encourage consistency in evaluation implementation.

Second, evaluation policies help make evaluation a more transparent and democratic endeavor. They constitute a public stance that an organization takes regarding evaluation. Because they are public, written policies, they can be known by everyone in the organization and thus criticized and challenged. Participation and dialogue can occur about which policies make the most sense under which circumstances.

Third, evaluation policy is also a mechanism for broader learning about evaluation. Preskill (2007) emphasized the value of viewing evaluation as a type of learning. Evaluation policy could be one of the key mechanisms for

such learning. Why? Because if we write evaluation policies down, we can archive them. We can share them. We can look at which of them seem to work better in which situations. That is, there can be some cumulative knowledge about what kinds of policies appear to work under various circumstances. We can also use evaluation policy to learn about the connection between evaluation theory and practice. Systematic reviews of evaluation implementation across organizations and settings can compare evaluation policies with respect to feasibility, effectiveness, efficiency, and other characteristics.

Fourth, evaluation policy is potentially an efficient mechanism for changing practice. For instance, if you want to change an evaluation practice in a large organization (e.g., the government), you typically have to go to each specific context and make the change locally. But if evaluation policies affect evaluation practice, the easier and more efficient way would be to change the overall policy once and have that change cascade to all relevant practice subcontexts.

Finally, evaluation policy is important because many of the controversies in evaluation today are essentially about such policy. This is apparent in the debates about the evaluation requirements in the Program Assessment Rating Tool (PART) system of the U.S. Office of Management and Budget (AEA Evaluation Policy Task Force, 2008), where the randomized controlled trial is presented in policy guidance as a preferred methodology for effectiveness evaluation in all programs in the federal government. We encountered similar issues in this association several years ago around regulations—essentially evaluation policies—that the Department of Education was proposing with respect to randomized experiments (Julnes & Rog, 2007; U.S. Department of Education, 2003). Controversies of this kind are fundamentally issues about what evaluation policy should be. There is no ignoring the issue of evaluation policy; it keeps coming back to us.

Who Controls Evaluation Policy? There are several fundamental issues related to power and participation that are central to the evaluation policy development process. Certainly, one of the major questions is who should be involved in formulating policies. The short answer is that evaluation policies will generally be best when there is broad and open involvement in policy development. Participatory approaches help ensure both that important issues don't get overlooked and that there is clearer communication and buy-in about evaluation.

Following closely on the question of who participates is *how* they can be involved most effectively. The field of evaluation has considerable shared experience with encouraging and facilitating participatory collaborative methods (Cousins & Whitmore, 1998; King, 1998), and we can effectively leverage this experience not just for evaluation itself but also in developing evaluation policy.

Questions of power and control are absolutely central to the discussion of evaluation policy. People in a position of power often have primary

control over policy making. But it is not clear that policies developed centrally will be appropriate or effective across an organization. Collaboration of stakeholders from throughout the organization will almost surely help encourage development of policies that are more appropriate and feasible.

Finally, no consideration of power and control in evaluation policy would be complete without raising issues related to fears and anxieties about evaluation generally and how they affect formulation of evaluation policy specifically. For instance, we know that many people are concerned about evaluation because they fear, often legitimately, that it may threaten their interests, harm their favorite programs, lead to decisions they don't like, take too much time or too many resources, and so on. Such fears and anxieties inevitably will and should play out in the evaluation policy development process. People will want to be assured that evaluation policies are fair and ethical, work for their perceived interests, and do not impose burdens from their point of view. To the extent that they can control the policy-making process, people will certainly attempt to do so. Evaluation policy development needs to be able to balance the interests of many stakeholders.

This discussion of power and participation does not begin to do justice to the complexities and challenges that must be addressed. We will need good evaluation work about the evaluation policy-making process in order to begin to understand how these issues unfold and what their implications might be. While we await the evolution of such work, and encourage its undertaking, we turn our attention from the broad definitional and conceptual issues to ones that are more methodological in nature.

An Evaluation Policy Methodology

If we are to have effective evaluation policies, we will need to develop methods for creating, archiving, evaluating, and revising them. This section attempts to move us along methodologically by offering an evaluation policy model that includes a taxonomy of evaluation policy types, a structure that depicts the interrelationships between policies and practices, and a set of principles that can be used in developing and managing policies.

Several caveats are in order before venturing onto this ground. First, the methodological framework offered here is only an initial, suggestive one. It has not itself been evaluated, although it is hoped that if it appears promising it will be subjected to extensive testing in practice. Second, the examples that are given here are for the most part created for purposes of this explication. There is a paucity of good evaluation policy examples and relatively few that were identified for this work. Readers should keep this in mind when interpreting these examples and should not use them as models or exemplars of policies recommended for real-world contexts.

Evaluation Policy Model: A Taxonomy. Evaluations are complex endeavors that involve many variations, dimensions, and activities. Commensurately, there will need to be many types of evaluation policies in order

to guide this complex endeavor well. Currently we do not have a clear taxonomy of the types of evaluation policies that might make up a comprehensive set of policies. Here is a tentative taxonomy, a list of evaluation policy categories within which we might wish to develop specific policies:

A Draft Taxonomy of Evaluation Policies

- Evaluation goals policies
- Evaluation participation policies
- Evaluation capacity building policies
- Evaluation management policies
- Evaluation roles policies
- Evaluation process and methods policies
- Evaluation use policies
- Evaluation of evaluation (meta-evaluation) policies

It would be preferable to develop a classification system for evaluation policies empirically. For example, one might use an evaluation method such as concept mapping (Kane & Trochim, 2006; Trochim, 1989; Trochim & Kane, 2005) to identify empirically what evaluators think the categories of policy might best be. Such efforts are currently under way, but until the results are in we can only anticipate what they might yield with intuitively based a priori categorizations such as those in the list. There is no single accepted form for expressing evaluation policies. So we are essentially free to experiment with different forms as this field evolves. To illustrate how such policies might look, the examples offered in the taxonomy are expressed in a somewhat formalistic policy language of the type that you might see in standard organizational policy and procedures manuals. Generally, evaluation policies should be relatively short and concise statements, although they may need to be supplemented with notes, definitions, and other explanatory text.

The taxonomy begins with policies that describe the *goals* of evaluation in the organization or context. For example, a general goal policy might be "The primary goals of evaluation in our organization are to learn about and improve programs and to ensure accountability to our stakeholders." One can imagine further subpolicies for each of the two goals in the broad policy. For example, you might describe the program improvement goal further with "Evaluation for program improvement should address both the process and outcomes of the program." A second evaluation policy category involves *participation* in evaluation and could address how and when stakeholders are to be involved. For instance, "Evaluations will be designed with input and consultation from key program stakeholders." Three categories of policies—on capacity building, management, and roles—are related to the organizational management, resources, and infrastructure that support evaluation. A high-level *capacity building* policy might simply state "The organization will

develop sufficient organizationwide capacity to support evaluation activities," with subpolicies providing more detail on what that means, such as "Staff will be trained in the methodology and use of evaluation appropriate to their program roles." *Management* policies might include policies that guide time resources ("Staff will be given sufficient time to accomplish evaluation-related activities"), policies that guide resource allocation ("Programs will allocate between 3% and 5% of total program costs for evaluation activities"), or ones that govern scheduling of evaluations ("All programs will conduct evaluations annually"). *Role* policies can be used to described the responsibilities that different people have for evaluation: "Program managers are responsible for ensuring that appropriate evaluations are conducted for each of their programs" or "Program staff are responsible for participating actively in integrating evaluation and program activities."

In the draft taxonomy all evaluation process and methods policies are collapsed into one category. To those of us who regularly work in evaluation, this may seem like too small a mechanism to contain policies that cover much of the detail about how we do our work, including question identification, sampling, measurement, design, and analysis. Perhaps so. It may be that we need several categories. Or we might subdivide this category as needed into several policy subcategories. Evaluators will see many such subdivisions, but for most nonevaluators this would be more detail than is helpful in a general taxonomy and so the broad category was used here. A *process and methods* policy might be something like "Wherever feasible, mixed methods (qualitative and quantitative) will be used in evaluations" or "Evaluation data will be stored in a secure location for no less than five years following the issuing of a report." The taxonomy includes a category for evaluation *utilization*, for instance: "Every evaluation must include a written plan for how results will be reported and used" or "Every evaluation should include an assessment of the utility of the results and recommendations for subsequent evaluations that eliminate approaches with low utility." Finally, the taxonomy includes a category for the evaluation of evaluation, or *meta-evaluation*, which could include "At least every three years the organization will contract for an independent meta-evaluation to assess the implementation, quality, and utility of its evaluations."

Notice that policies differ in their generality or specificity. An example of a broad or general "covering" policy in the process and methods category might be "Evaluations will use the highest-quality and most cost-efficient approaches and methods appropriate to the development of programs." Such general policies are essential, especially for their public and transparent signaling value, but they are not likely by themselves to be sufficient guidance for evaluation practice. They will almost always require development of subpolicies to articulate what they intend. Or another: "The highest professional standards will be used to ensure the rights and protections of evaluation participants." Notice the language here. The policy itself, at least as stated at this level, does not define what is meant by "the highest
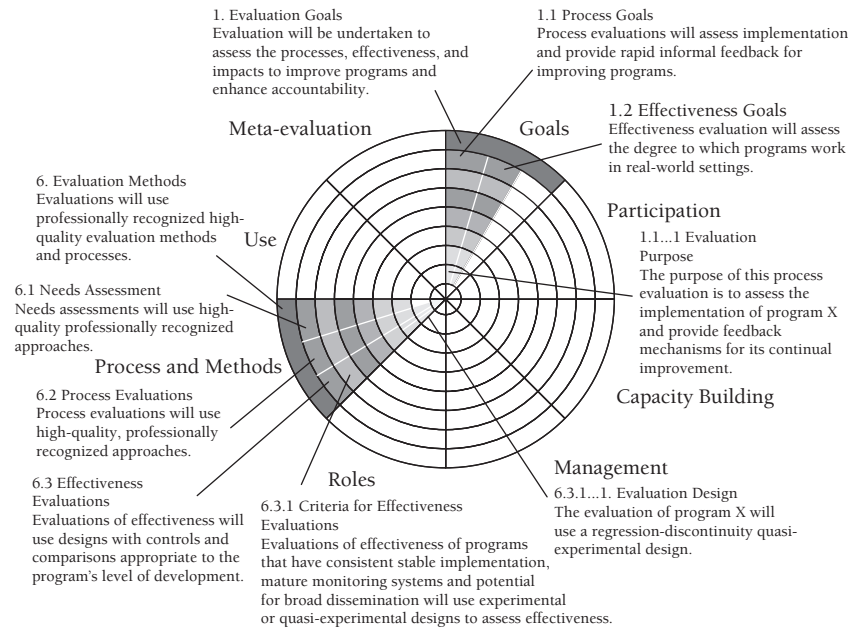
Figure 2.2. The Evaluation Policy Wheel
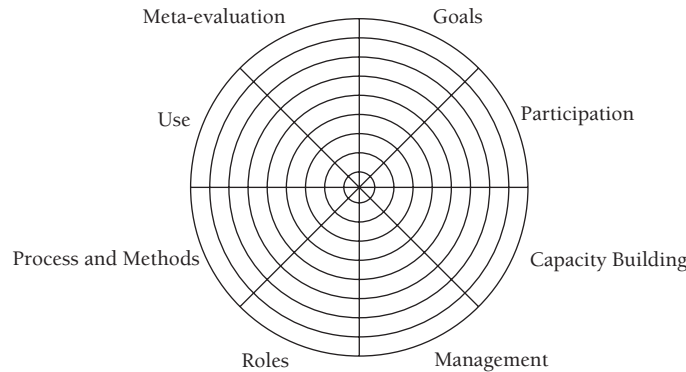
professional standards." Evaluators and other professionals have long debated what the highest professional standards are, and these are likely to evolve over time. But from a policy point of view this general covering policy plays a critical role in that it signals that organizations and units coming under the purview of this policy statement are delegated responsibility for defending how they translate or interpret the phrase "highest professional standards." Finally, it is important to note that a number of policies could legitimately be placed in multiple categories in the taxonomy. The last example could, for instance, be placed in the goals or the participants' policy category. It matters less where a particular policy is located in the taxonomy than the entire set of policies address all relevant aspects of evaluation.

Evaluation Policy Model: Structure. Our methodological framework needs a superstructure that can be used to manage complex multiple evaluation policies and their relationship to practice. The mechanism offered here is what we will call a "policy wheel," a visual rubric for policy development and evolution.[1] The policy wheel describes evaluation policy within a simple circle diagram (Figure 2.2). All evaluation policies can be located somewhere on the wheel. There are layers on the wheel, from the outermost area to the center of the circle, meant to suggest levels of generality of policy. The most general policies within their respective taxonomic categories are on the outer layers; more specific subpolicies and subsubpolicies are included in progressively inner layers. As you move to the center of the wheel, you transition from policies into practice, procedure, or operationalization. So the center of the circle is practice, and the outer layer is the highest-level policy. In effect, when you specify policies at such a level of detail that they leave no room for discretion or judgment, you are essentially prescribing the operational procedures that need to be followed.

The policy wheel is divided into eight wedges that correspond to the proposed types of evaluation policy in the taxonomy described earlier. To illustrate how the wheel is used, let's consider some hypothetical policies

Figure 2.3. Examples of Evaluation Goals and Evaluation Process and Methods Policies

1. Evaluation Goals
Evaluation will be undertaken to assess the processes, effectiveness, and impacts to improve programs and enhance accountability.

1.1 Process Goals
Process evaluations will assess implementation and provide rapid informal feedback for improving programs.

1.2 Effectiveness Goals
Effectiveness evaluation will assess the degree to which programs work in real-world settings.

6. Evaluation Methods
Evaluations will use professionally recognized high-quality evaluation methods and processes.

6.1 Needs Assessment
Needs assessments will use high-quality professionally recognized approaches.

6.2 Process Evaluations
Process evaluations will use high-quality, professionally recognized approaches.

6.3 Effectiveness Evaluations
Evaluations of effectiveness will use designs with controls and comparisons appropriate to the program's level of development.

6.3.1 Criteria for Effectiveness Evaluations
Evaluations of effectiveness of programs that have consistent stable implementation, mature monitoring systems and potential for broad dissemination will use experimental or quasi-experimental designs to assess effectiveness.

1.1...1 Evaluation Purpose
The purpose of this process evaluation is to assess the implementation of program X and provide feedback mechanisms for its continual improvement.

6.3.1...1. Evaluation Design
The evaluation of program X will use a regression-discontinuity quasi-experimental design.

Meta-evaluation

Goals

Use

Participation

Process and Methods

Capacity Building

Roles

Management

and how they might be arrayed. These hypothetical example policies are intended to look like real policies so that we can use them to illustrate the workings of the policy methodology.
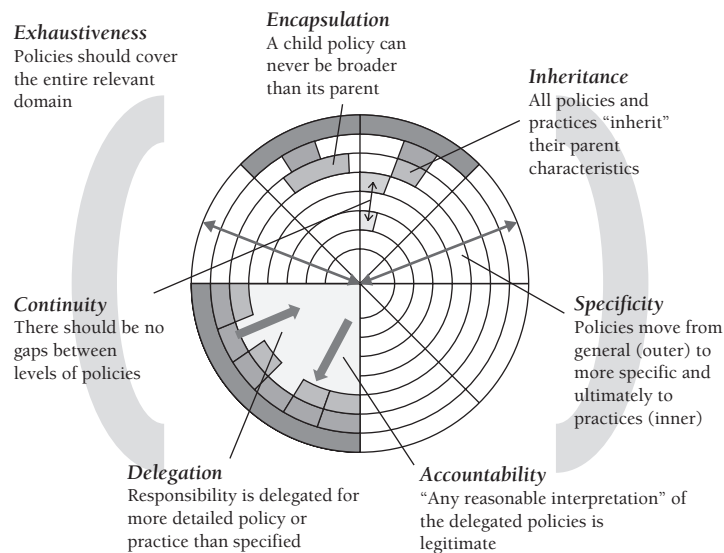
Let's begin by focusing on evaluation goals policies and consider a quick example. In the upper right portion of Figure 2.3 we see the goals policy wedge partially filled in with a hierarchy of potential evaluation policies in that area. We begin in the outer layer with the most general goal policy: "Evaluation will be undertaken to assess processes, effectiveness, and impacts to improve programs and enhance accountability." This is a very broad policy statement. There are any number of ways to make it more specific. We might want to start by specifying a more detailed policy for one aspect of this general goal, the topic of process evaluation: "Process evaluation will assess implementation and provide rapid informal feedback for improving programs." We could do more, and specify a comparable policy for effectiveness evaluation as illustrated. If we keep defining what we mean more specifically by each successive layer of policy, eventually we are going to get down to something that essentially is a description of what we will do in practice. For example, in this case we actually get to something that is essentially a description of the purpose of a process evaluation: "The purpose of this process evaluation is to assess the implementation of program X and provide feedback mechanisms for its continuous improvement."

Note that we can keep track of the policy hierarchy by numbering the policies. Because the goals policy type is the first in our taxonomy, we would number the highest-level policy in this area with a 1, the next layer of subpolicies 1.1, 1.2, and so on, until we get to something like 1.1 . . . 1 or 1.1 . . . 2 for the most detailed policies. This means we can alternatively list all of the policies in a hierarchical outline, a form that will be easier for reading them. However, the wheel is a useful visual display because it shows at a glance how specific or general the policies are across the entire taxonomy.

Consider another set of hypothetical example policies, this time in the area of process and methods policies. It begins with a general policy requiring "high-quality" evaluation methods and processes, differentiates several types of evaluation (needs assessment, process evaluation, and effectiveness evaluation), and details what constitutes high-quality methods and processes for these evaluation types. At its most specific, the example policy is essentially a prescription for what should be done under a particular set of circumstances to accomplish effectiveness evaluation—in this hypothetical example, a regression-discontinuity design, of course (Trochim, 1984)!

Evaluation Policy Model: Principles. We can also use the policy wheel as a framework for illustrating the major principles that guide the evaluation policy model (Figure 2.4). We have already seen that policies change in specificity as we move from the outer levels to the inner. This is referred to here as the principle of *specificity*. The notion that subpolicies and

Figure 2.4. Principles of the Evaluation Policy Model

Exhaustiveness
Policies should cover the entire relevant domain

Encapsulation
A child policy can never be broader than its parent

Inheritance
All policies and practices "inherit" their parent characteristics

Continuity
There should be no gaps between levels of policies

Specificity
Policies move from general (outer) to more specific and ultimately to practices (inner)

Delegation
Responsibility is delegated for more detailed policy or practice than specified

Accountability
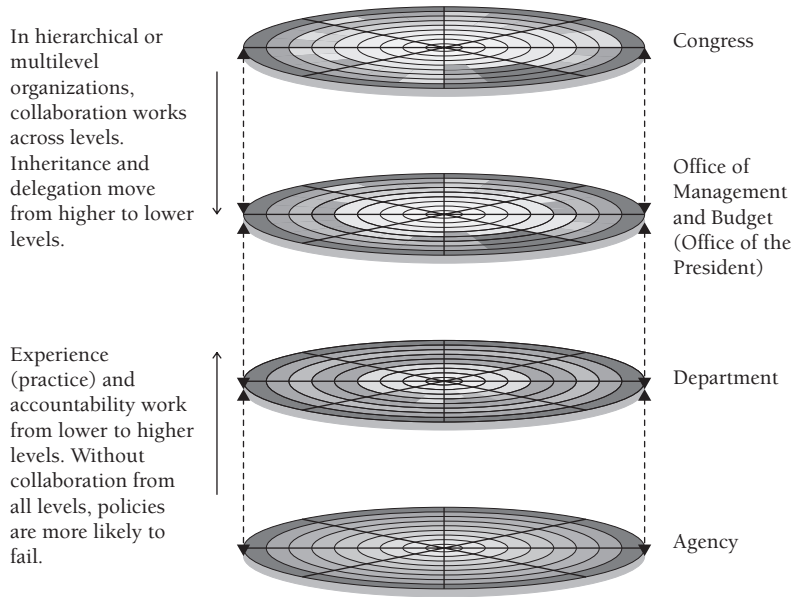"Any reasonable interpretation" of the delegated policies is legitimate

ultimately practices inherit their outer-level parent policies is a principle or characteristic that can be labeled *inheritance*. That is, when you operationalize a set of policies in practice it's expected that you will enact all of the policies in the hierarchy, not just the most specifically stated one. The idea that the policy-practice continuum is hierarchical—that broader policies contain within them subpolicies, which contain within them further subpolicies—is what can be termed the principle of *encapsulation*. Policies should cover the entire relevant domain in their taxonomy and not omit anything important, a principle we might call *exhaustiveness*. There should be continuousness of specification from general policy to operational practice, or what we might call the principle of *continuity*. For instance, imagine an agency that has only one very high-level evaluation policy such as "We should have regular evaluations of all major programs" and one specific policy (almost at the level of procedure) such as "We should evaluate with randomized experiments every other Tuesday." In this hypothetical case, we are essentially jumping from a very high-level policy into something that is quite specific. Many of the controversies we encounter in real-world evaluation policy are related to this type of jump, to the lack of policy continuity. The principle of *delegation* is the idea that policies are developed to a certain level of specificity at one hierarchical level in an organization and to other levels of specificity at others. For instance, in a large, complex, multilevel organization, the broadest-level policies tend to be developed at the highest level of the hierarchy. Lower levels are delegated the responsibility of developing more specific policies appropriate for their level, with the operational staff ultimately delegated responsibility for translating policies into actual practice. Where delegation occurs, those delegated to are responsible for defending the judgments they make about how they operationalized policies, a principle we label *accountability*.

This principle of accountability is a challenging one and warrants a bit more explanation. It is based on the idea of "any reasonable interpretation" (Carver & Carver, 2006). Typically, evaluation policies are not and would not be fully specified from general policy all the way to specific practice. That is, in the real world we would never expect the policy wheel to be completely filled in. In fact, in some organizations it might be perfectly reasonable to specify only the highest-level or outer-circle policies in each area of the taxonomy and delegate responsibility to staff to determine appropriate ways to implement them. When policies are only specified to a certain level, the assumption of this model is that "any reasonable interpretation" of the most specifically stated policies would be acceptable. If you specify only a high-level process and method policy such as "Evaluations will use professionally recognized, high-quality evaluation methods and processes" and leave it at that, then any evaluation done that can be considered a reasonable interpretation of the policy should be accepted by the policy makers. Note that this puts the burden of articulating policies to the level of specificity desired on whichever person or group is responsible for policy. If an

evaluation that is done meets the criterion of a reasonable interpretation of the policy but is not satisfactory to those responsible for the policy, this suggests that they need to specify more detailed levels of policy in this area. This illustrates the dynamic tension among delegation, accountability, and specificity. The less specific your policies (i.e., the more you rely on a few general policies), the more you are delegating to others the responsibility for deciding what constitutes a reasonable interpretation. Although with greater delegation staff members have greater flexibility, they also have more accountability to demonstrate that their interpretation was a reasonable one. When policies are specified in greater detail, delegation and flexibility are by definition reduced, but so is the burden or challenge of accountability.

In complex organizations or systems it is likely that policy making will itself be a hierarchical process with different levels of the hierarchical organization developing different levels of policy. High-level policies (outer circles) would be specified at the highest level of the organization, and responsibility would be delegated to subunits that inherit with those policies the responsibility for determining more detailed policies and ultimately evaluation practice. For example, consider how this might play out in the context of the U.S. federal government (Figure 2.5). At the highest level in the policy organization hierarchy—in the case of the U.S. government, this would be the Congress, which is responsible for making law—they would be expected to specify very high-level evaluation policies. The next level in

Figure 2.5. Policy in a Hierarchical Organization or System

In hierarchical or multilevel organizations, collaboration works across levels. Inheritance and delegation move from higher to lower levels.

Congress

Office of Management and Budget (Office of the President)

Experience (practice) and accountability work from lower to higher levels. Without collaboration from all levels, policies are more likely to fail.

Department

Agency

the policy hierarchy might be in the Executive Branch, for instance, at the level of the Office of Management and Budget in the Office of the President. They would inherit the policies that Congress set and add more specific policies of their own. With each move further down in the hierarchy to the cabinet-level department, agency, subagency, and so on, more detail may be added into the policy wheel. At the lowest level, the "policies" are essentially short descriptions of the procedures that will be followed in conducting evaluations.

It would not be desirable for this multilevel process to be unidirectional. Certainly, policy should guide practice. But the reverse is also true: policy needs to be informed by practice. That is, we need an experiential and empirical practice base to inform policy development. Without that experiential base, policies are more likely to be ill-suited for the context and more likely to fail.

Real life seldom follows a rational model like the one offered here. For instance, if you are in a lower level of an organizational or system hierarchy (and aren't most of us?), you don't need to wait for the higher levels to formulate policy before you can develop policies at your level. Depending on where you are in the organizational structure, you can enter the policy development process at any level of this model. When you enter the hierarchy at some level and survey the existing policy landscape, you'll see that you already likely inherit a number of written or unwritten policies. And policies you help develop will likely have implications for lower levels of the hierarchy, either now or in the future.

Evaluation Policy Challenges and Opportunities

Challenges. Developing comprehensive written evaluation policies is not an easy endeavor; there will be a number of important challenges and barriers that have need to be addressed as this area develops. For instance, there is a real danger that in our zeal to articulate sensible policies we will *overformalize* the effort and develop an ineffective bureaucratized approach that actually stifles good policy making and becomes a detriment to evaluation. Finding the right balance between which policies would be specified explicitly in written form and which would evolve as implicit norms is going to be a major challenge.

Everyone who's ever worked in a large organization knows that *senseless or illogical policies* are legendary and generate a significant amount of the griping that occurs. We need research and evaluation on the effectiveness of different approaches to developing evaluation policy that help us understand how we might best avoid senseless and contradictory policies.

One of the most frustrating aspects of much organizational policy is the tendency for well-intended policies to have *unintended negative side-effects*. Unintended negative effects of policies are often the result of systems challenges; one part of the system establishes a policy that makes perfect sense

to them without thinking through or realizing the negative effects it will have on another part of the system. We need to develop ways to rapidly evaluate new policies to determine when unanticipated negative effects might result.

Perhaps one of the most important challenges we are likely to face is that of *ethics*. Many of the ethical issues involved in evaluation policy have already been mentioned, including the implications of power and control issues, the importance of transparency, and the need to engage all relevant stakeholders in the policy formulation process. Evaluation policy making must be guided by our core values.

There are undoubtedly many other challenges and barriers to good evaluation policy making than just these. We will need good evaluation work to identify and catalogue such challenges and even better work to assess how best to address them.

Opportunities. It is an old adage that "with challenge comes opportunity," and this certainly is the case in the area of evaluation policy. Here are a few potential ideas evaluators might explore as we move ahead in this arena; this list might also be viewed as an agenda or a set of "next steps" to be pursued.

Evaluation Policy Audits. We desperately need descriptive studies of current evaluation policies, and this would be something we could get to work on immediately. Take a look at the evaluations you are working on and the organizations in which they are set. Determine existing evaluation practices and assess which are based on explicit written policies and which are more implicit. Conduct a systematic review of documents, including reports (such as annual reports, project reports to funders) and internal memos and communications, to identify existing evaluation policies. Relate the policies to a taxonomy, either one that already exists or one developed in the local context. Such taxonomies can in turn be used as checklists for auditing the as-is situation and suggesting potential gaps. We need to develop taxonomies and corresponding audit methodologies that are better than the ones suggested here, and these need to be informed by systematic research. For instance, research that is currently under way uses concept mapping with staff in a complex multicounty cooperative extension organization in New York State to develop a taxonomy of evaluation policies and corresponding audit procedures that would be empirically grounded (Hargraves & Trochim, 2008) and contextually appropriate.

Evaluation Working Groups for Policy Development. Very often the impetus for developing evaluation policies will arise when you are in the midst of conducting an evaluation and it becomes apparent that there is an absence of locally adopted written policies to guide your work. In such circumstances, an evaluation working group or evaluation advisory group might be a useful mechanism for initiating and facilitating an evaluation policy development effort in conjunction with ongoing

evaluation. The group could be responsible for overseeing a policy audit and making recommendations for policies that need to be developed. The policies that result can then be shared with others in their organization and might form the basis of a more organizationwide subsequent policy effort.

Evaluation Policy Clearinghouses and Archives. As we begin to accomplish evaluation policy audits, it would be valuable to initiate a systematic effort to format them uniformly, archive them, and make them available publicly. Archives and clearinghouses will both encourage greater communication about policies and yield a database that would be invaluable for research, evaluation, and systematic review and synthesis, helping the field identify the most effective policies in various organizations and conditions. That is, systematic archiving would help ensure that we can be cumulative about what we learn about evaluation policy.

Evaluation Policy Analysis. More formally, we should be developing the idea of evaluation policy analysis, based on notions of the traditional field of policy analysis. One key role of such an effort would be to conduct systematic empirical comparisons of how alternative policies work and under what circumstances. If we had an archive of evaluation policies across multiple organizations, it would be possible to determine the range and types of policies in any category of the taxonomy. For instance, one of the most common questions raised by policy makers about evaluation is how much of their program budget they should allocate for evaluation. Evaluation policy analysis could look across multiple sets of policies, determine their allocations, assess the variation across contexts, and even relate them empirically to subsequent quality and efficiency of evaluations conducted.

Software Systems for Evaluation Policy Management. There is a need to develop information and software technology that can be used to help groups and organizations manage evaluation policies. For instance, we might imagine something like a Web-based software tool based on the hierarchical policy wheel of Figure 2.2. It could give complex organizations a platform for collaboratively entering and harmonizing evaluation policies throughout their systems. It could enable users to toggle back and forth between a graphic representation of policies on the wheel and a hierarchical outline of such policies. This would be an especially useful tool for very large, complex organizations where each level of the organization could enter its policies and print what is appropriate at its level while also being able to see whether there are gaps in policy specification or how their policies compare with what other groups in the organization are doing.

Consulting on Evaluation Policy. There's a potential cottage industry out there for evaluation consultants in connection with evaluation policy. One can imagine evaluation policy consultants who are contracted by organizations to facilitate development and management of their evaluation policies,

who would be responsible for managing policy audits, reviewing potential policies from other similar organizations, archiving the policies that are developed, and so on. Here's where evaluators' extensive knowledge of both the field of evaluation and of participatory and facilitative methods could be put to good use to enhance evaluation policy.

Harmonizing Evaluation and Other Policies. It is important to recognize that evaluation does not exist in a vacuum. Although this chapter has focused on evaluation policy, most organizations already have (sometimes extensive) written policies to cover everything from personnel management to acquisitions and contracting, to strategic planning and finance. Evaluation policies need to be integrated into the larger organizational policy framework that already likely exists. This means that evaluators must become more familiar with the methodologies and structures for developing traditional policy and procedure manuals in organizations and ensure that our work can be integrated appropriately.

Conclusions. A central message of this chapter is that we have to be more serious about evaluation policy. We have to encourage more written, public, and transparent evaluation policies. We have to develop a balance between the desire for general policies that cover an organization consistently and the need for flexibility and adaptation. We have to address and acknowledge the issues of power, hierarchy, delegation, and incentive; encourage collaboration and participation; and encourage archiving and sharing of policies.

Perhaps it's fitting to conclude by recalling that evaluation policy is important because it guides evaluation practice, which in turn influences the quality of the programs and policies that shape people's lives. For many evaluators, providing compelling analysis of and practical suggestions for improving public policies and programs is one of the best ways we can enhance our society. In the United States we are at an especially important moment with the change in Washington to the Obama administration, which faces a national debt in the trillions of dollars, annual deficits in the hundreds of billions, and uncertainties about financial institutions and the economy. At the same time, concerns remain about national security, health care, education, energy development, and many other facets of American life (AEA Evaluation Policy Task Force, 2009).

To the Obama administration we need to say that program evaluation is essential to addressing these issues. It can help answer new questions about current and emerging problems, reduce wasteful spending, increase accountability, and support major decisions about program reforms and improvements. President Obama has pledged to review all government programs to identify those that work and those that don't, and to make vital programs work better than they do now. We need to encourage the new administration to examine not just government programs but also how it evaluates them. We need to help key political appointees and

senior careerists in the Executive Branch and the Congress to make program evaluation integral to managing government programs at all stages, from initial development through startup, ongoing implementation, and reauthorization. A more concerted and coherent effort to develop sensible evaluation policies, to archive and make them available to others, and to conduct evaluations to assess their effectiveness is critically important to making an effective case for evaluation in the Obama administration and in the future.

Note

1. Several key features of this model are adapted from the work of John and Miriam Carver on the governance policies for boards of directors of organizations (Carver, 2006; Carver & Carver, 2006). The unique needs of evaluation policy required adaptations and considerable extensions to these features, which undoubtedly distort or change them, hopefully appropriately for this context. The methodology offered here owes a clear debt to their work, but the Carvers should bear none of the responsibility for the faults of these adaptations.

References

AEA Evaluation Policy Task Force. (2008). *Comments on "What constitutes strong evidence of a program's effectiveness?"* Retrieved March 19, 2009, from http://www.eval.org/aea08.omb.guidance.responseF.pdf

AEA Evaluation Policy Task Force. (2009). *An evaluation roadmap for a more effective government.* Retrieved March 19, 2009, from http://www.eval.org/aea09.eptf.eval.roadmapF.pdf

Carver, J. (2006). *Boards that make a difference.* San Francisco: Jossey-Bass.

Carver, J., & Carver, M. (2006). *Reinventing your board: A step-by-step guide to implementing policy governance.* San Francisco: Jossey-Bass.

Chen, H., & Rossi, P. (1990). *Theory-driven evaluations.* Thousand Oaks, CA: Sage.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80*, 5–23.

Fetterman, D., Kaftarian, S. J., & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability.* Thousand Oaks, CA: Sage.

Hargraves, M., & Trochim, W. (2008, November). *Designing and implementing evaluation policies to sustain evaluation practice in extension programs.* Paper presented at the Annual Conference of the American Evaluation Association, Denver, Colorado.

Julnes, G., & Rog, D. (Eds.). (2007). Informing federal policies on evaluation methodology: Building the evidence base for method choice in government sponsored evaluation. *New Directions in Evaluation, 113.*

Kane, M., & Trochim, W. (2006). *Concept mapping for planning and evaluation.* Thousand Oaks, CA: Sage.

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation, 80*, 57–67.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). London: Sage.

Preskill, H. (2007, November). *Evaluation's second act: A spotlight on learning.* Paper presented at the 2007 Annual Conference of the American Evaluation Association, Denver, Colorado.

Trochim, W. (1984). *Research design for program evaluation: The regression-discontinuity approach.* Thousand Oaks, CA: Sage.

Trochim, W. (Ed.). (1989). *Concept mapping for evaluation and planning* (Vol. 12, 1). New York: Pergamon.

Trochim, W., & Kane, M. (2005). Concept mapping: An introduction to structured conceptualization in health care. *International Journal for Quality in Health Care, 17*(3), 187–191.

U.S. Department of Education. (2003, November 4). Scientifically based evaluation methods. *Federal Register*, 62445–62447.

WILLIAM M. K. TROCHIM is professor of policy analysis and management at Cornell University and the director of evaluation for the Weill Cornell Clinical and Translational Science Center, the director of evaluation for extension and outreach, and the director of the Cornell Office for Research on Evaluation.