



Cornell University

# Evaluation Policy and Evaluation Practice

---

Presidential Address  
2008 Annual Conference  
of the  
American Evaluation Association

William M.K. Trochim

Do not use, quote or cite without written permission of author.

### **Author's Note**

This document is a loose and heavily-edited transcription of the Presidential Address that I delivered on November 6, 2008 at the Annual Conference of the American Evaluation Association in Denver, Colorado. While it stays true to the speech as delivered, I made a number of stylistic edits to deal with the problem of getting a speech to read well. In the address itself I wound up pressed for time and had to omit a considerable amount of content. In this written version I added some of the key points that I omitted, while also trying to stay true to the spirit of the speech. I am in the process of working with my two colleagues Leslie Cooksy and Mel Mark to develop a New Directions in Evaluation volume that will include a paper that covers the material in this speech but presents it in greater detail and, hopefully, in a better organized fashion.

### **Acknowledgements**

I want to thank Mel and Leslie for their support in developing the Presidential Strand for the conference and for their ongoing collaboration on this topic. I would also like to thank my colleagues at Cornell for helping me think through many issues covered in this address. I would also like to recognize the considerable contribution that Richard Gallagher made in helping me prepare earlier drafts of the speech and some of the slides. Finally, I wish to acknowledge my spouse, Mary Kane Trochim, who had to live with me through the agony of preparing this address.

### **Author Contact Information**

William M. Trochim  
Professor, Policy Analysis & Management  
Director, Cornell Office for Research on Evaluation (CORE)  
Director of Evaluation, Clinical & Translational Science Center  
Director of Evaluation for Extension & Outreach  
607-255-0887  
607-255-1150 fax  
[http://www.socialresearchmethods.net/  
wmt1@cornell.edu](http://www.socialresearchmethods.net/wmt1@cornell.edu)

Mailing Address:  
Martha VanRensselaer Hall  
Cornell University  
Ithaca, NY 14853

Physical address (deliveries and visitors):  
2nd Floor Mann Library

# Evaluation Policy and Evaluation Practice

William Trochim  
Cornell University

## Introduction

Today I would like to talk about evaluation policy and evaluation practice. Here's the kind of stuff I'm going to go through (Slide 2). I won't take a lot of time talking us through it. I'm going to talk about "What is evaluation policy?" and give you some idea of why it is important and how we can make effective policy. I'm not going to give a definitive description of evaluation policy, but I will give you some thoughts about how we might approach it. I certainly want to highlight for us some of the many, many challenges we are going to face as we talk about the topic of evaluation policy, and give some idea of "where do we go from here?" I have a few thoughts about that. I'm actually hoping that you all have thoughts about that and that this conference is going to be a time for us to think about the implications.

## Overview

---

- What is evaluation policy?
- Why is evaluation policy is important?
- How can effective evaluation policy be made?
  - Who's involved in making evaluation policy?
  - What approaches can we use?
  - What types of evaluation policies are there?
  - How is evaluation policy structured?
  - What are the major characteristics of evaluation policy?
- What are the major challenges in evaluation policy?
- Where do we go from here?

2



When I picked this topic of "Evaluation Policy and Evaluation Practice" as the theme for this conference last year, and tried it out on some of my friends I quickly realized that I was in some trouble because people immediately thought that what I was talking about is policy in general, not evaluation policy specifically. I spent the better part of the year trying to explain the distinction between policy and evaluation policy, but I want to start with that because I think for some it may still be a point of confusion.

## Major Policy Examples

---



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.

May 25, 1961

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.

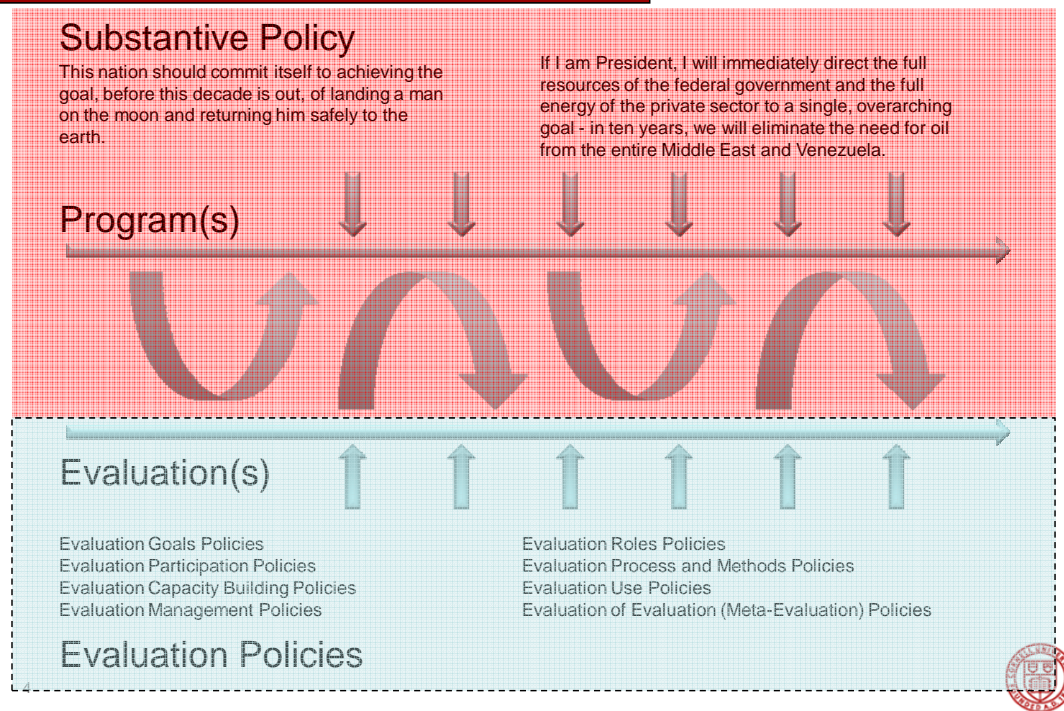
August 4, 2008



3

When I think about policy (Slide 3), I think as most people do about big policy, major national policy, 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.” It certainly had some of the major defining features of any policy. It had a definable outcome in mind. It had a time-frame in mind. Notice something very important about this kind of policy statement: John Kennedy didn’t talk about payloads, didn’t talk about lunar landers, didn’t talk here about astronauts drinking Tang. In fact, I don’t even know if they had Tang at that point. So, there is something important about that with respect to policy. This is to me an example of a substantive policy. Here’s another one, perhaps a little more recent: President Elect Barack Obama. You knew I had to find some way to sneak in a picture about this. So, in August, President Elect Barack Obama said, “If I am President, I will immediately direct the full resources of the federal government and the full energy of the private sector,” though I’m not quite sure how he can do that, “to a single, overarching goal - in ten years, we will eliminate the need for oil from the entire Middle-East and Venezuela.” Now, that’s not yet a policy because he is not yet president, but if he does get to be President and he declares that statement, then it becomes a policy, a very high-level policy. Notice that he doesn’t talk about the details of how he is going to get to that policy. It is really articulating the direction.

## Substantive versus Evaluation Policy

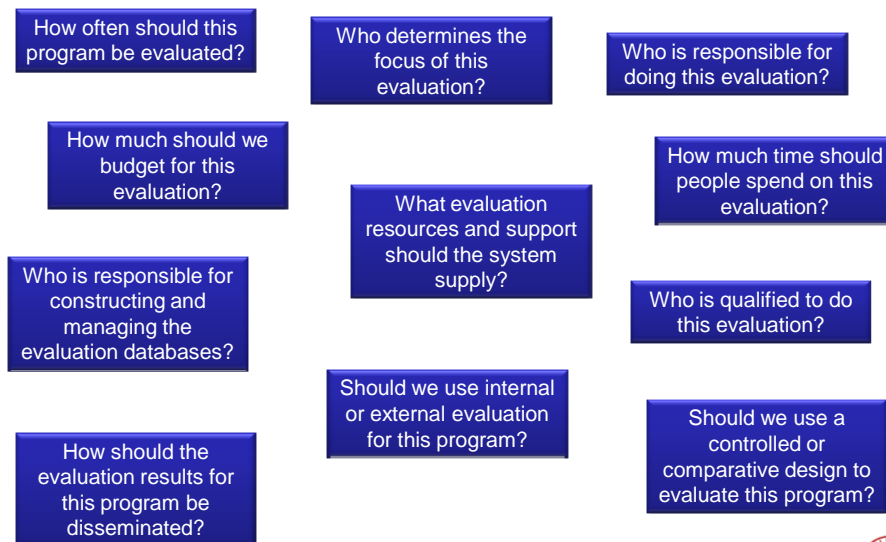


I want to make a distinction now then between what I will call here “Substantive Policy versus Evaluation Policy” (Slide 4). I was tempted to call this “*Public* Policy versus Evaluation Policy,” but there are policies in other than public agencies, and so I wanted to distinguish evaluation policy from any kind of substantive policy, public or otherwise. When I think of substantive policy I think of the kinds of examples of public policies I just gave. Usually policies like that get translated into things like lunar landers, Tang for astronauts, special new technologies for fuel efficiency, or so on. And, we might call what it gets translated into programs or practices or our activities. But there is a translation between policies and what actually happens. Evaluation, as we all know - and I won’t spend too much time on this - operates in parallel with anything that we might be doing programmatically. That is, evaluation is an ongoing endeavor that can provide us with essential feedback about what’s happening in the programs or practices or technologies associated with policies. There is a dynamic throughout the life of policy, or the programs or activities associated with that policy, and evaluation plays a critical role. Evaluation *policies*, in effect, guide how evaluation happens. There is something happening that I show at the bottom of the slide, a whole bunch of evaluation policies -- I’m previewing down here a little classification of them that I am going to come to in a little bit. So, we have evaluation policies that help to determine what kind of evaluation we are doing, and that evaluation is yoked to the policies and programs that are involved.

I would like to emphasize that the top part of this slide is not what the focus of my talk is going to be. What I really want to talk about today is what is happening down here on the bottom of the slide. That is, evaluation policy and evaluation practice. So, I apologize to those who thought we were going to be talking about going to the moon or becoming energy independent. Let’s not lose sight of the top of that slide because the reason we are doing evaluation is in

service of those things that we are evaluating - that is a critically important part of what we do. That is, I don't mean to demean the top part of this. That's really the central focus for why we are doing evaluation, but I do want to point out that in this talk I am going to go in a different direction and focus on the relationship of evaluation policies to evaluation and its practice.

## Evaluation Policy Questions Are *Everywhere*



5



Evaluation policy is everywhere (Slide 5). Take a look at some of these questions, and ask yourselves whether or not they are ones that you have encountered with respect to a project that you have done, or with respect to evaluations that you have managed (if you are an evaluation or policy manager). Have you ever been in a situation where the question has been “Who is responsible for doing the evaluation?” “Who is responsible for constructing and managing the evaluation databases?” I am running into that problem right now. In evaluation we’ve got complex databases that involve multiple stakeholders. How do we decide who should be responsible for collecting, for managing those databases, for setting them up? “How often should this program be evaluated?” Should we evaluate it every year? Does it even make sense to talk about a time-frame for an evaluation? Perhaps it should be ongoing? How do we do that? “Should we use controlled or comparative design to evaluate this program?” Is this program one that we should be assessing with a randomized experiment or quasi-experimental design? “Who determines the focus of this evaluation?” “How should the evaluation results for this program be disseminated?” “What evaluation resources and support should the system supply?” “How much should we budget for this evaluation?” One of the most frequent questions I get, and I have gotten this year as President of AEA, is what percentage of our budget should be spent on evaluation? “How much time should people spend on this evaluation?” That’s a really good question. There are lots of different people who may be engaged in evaluation, and the question of how much or what percentage of their time should they allocate to the task - five percent, fifteen percent, half a day a week, half a day a month? “Should we use internal or external evaluation for this program?” and what does that mean? What determines when we should use

external evaluation? “Who is qualified to do this evaluation?” My dear friend Deb Duran at NIH keeps asking me who is a qualified evaluator, and how are we going to know it. Do these questions sound familiar? I see all of these questions as evaluation policy questions. So, the issue I’m going to be talking about today is how do we develop policy in our work context that really helps us address these kinds of questions?

## **What is Evaluation Policy?**

---

An evaluation policy is...  
a rule, regulation, law, principle or norm  
that a group or organization **uses**  
to guide **decisions and actions**  
regarding **evaluation**

*All* evaluations are *already* guided  
by evaluation policies

Most are “informal” and unwritten

6



Well, you know I had to try to give a definition (Slide 6). I hate definitions because no definition is going to ever be perfect, but here’s a good a good start I think. I’m saying that an evaluation policy is - and notice how I am hedging here on the top line - a rule, a regulation, a law, a principle or a norm. I mean, I may as well throw in a kitchen sink, right? But evaluation policy is that broad. It can manifest itself in many ways (I have my own little approach to thinking about that, as I will show you shortly). I will also say that evaluation policy is something that a group or organization uses. Although an individual can have a policy for himself, generally speaking, policies are across individuals. In effect, evaluation policies are systems issues. They are used to guide decisions and actions. If we just stop there - “an evaluation policy is a rule, regulation, law, principle or norm that a group or organization uses to guide decisions and actions” - we have a pretty good definition of policy in general. Now, what makes something an *evaluation* policy? Well, it’s that “regarding evaluation” part of the definition.

Here is a point that I want to make that may or may not survive this talk, but I would like to try it anyway. I want to make the case that all evaluations are *already* guided by policy, whether you realize it or not. The problem is that most of the policies that guide evaluations are informal. You would have a hard time pointing to them and they are generally not written down. So, a lot of my case in this talk is going to be encouraging us to find ways to make what is informal, implicit, or not written down a little bit more formal and explicit.



## What is Evaluation Policy?

### How do we distinguish policies from guidelines, standards or theories?



*"Are these just guidelines, or are they actual new policies?"*



One of the key definitional questions about policy has to do with how we distinguish it from other things that sound like it (Slide 7). How do we tell the difference between a policy and a guideline, a policy and a standard, or a policy and a theory? I actually have what I think is a pretty simple operational definition that will distinguish evaluation policy from these other terms. First, we have to recognize that a standard or a guideline or a theory is not a policy. But each of these can become a policy when we decide to use them in some context. So, if you have guidelines, such as the AEA Guiding Principles for Evaluators (<http://www.eval.org/Publications/GuidingPrinciples.asp>), these are what they are -- just general guidelines. They become evaluation policy when a group or organization says they are going to *use* or adopt these guidelines for making decisions about evaluation. If we have theory about how to do evaluation, such as the theory that underlies utilization-based evaluation, empowerment-evaluation, theory-driven evaluation, or any other form of evaluation, those theories are simply what they are -- theories. They become the foundation for evaluation *policy* when an organization says, "We are going to use this approach or we are going to adopt this method." What is really interesting here is that policy always has a connection to the real-world, always has a connection to practice. We can see this distinction reflected well in the cartoon where a person is raising his hand and asking "are these just guidelines or are they actual new policies?" In effect, what he is really asking is "is this something we are going to use or are these just general guidelines?" In other words, he wants to know if we're serious about this or not. And, he wants to know if this is the public or official stance of the organization or not. I think this notion of "use" and the emphasis on the public signal about use is a good way to get at the distinction between things like guidelines, standards, and theories on the one hand and policies on the other.



## What do Evaluation Policies look like?

All programs will conduct annual program evaluations that address both implementation and effectiveness.

Staff will be provided sufficient time to accomplish evaluation-related activities

The highest professional standards will be utilized to assure the rights and protections of evaluation participants.

Programs will allocate between 3-5% of total program costs for evaluation activities.

Staff will be provided training in the methodology and use of evaluation appropriate to their program roles.

The organization will develop sufficient organization-wide capacity to support evaluation activities.

Evaluations will be designed with input and consultation of key program stakeholders.

Evaluations will utilize the highest quality and most cost-efficient approaches and methods appropriate to the development of programs.

8



What do evaluation policies look like? I'm just going to quickly go through some examples (Slide 8). Generally, evaluation policies are relatively short statements. In a comprehensive set of policies we are likely to group individual policy statements into collections around different topical aspects evaluation. So, here is one example evaluation policy (By the way, I made these up, so I'm not necessarily recommending you adopt these example policies. In real-world contexts we have to work together to figure out what policies make sense.): "Programs will allocate between 3-5% of total program costs for evaluation activities." Here's another: "All programs will conduct annual program evaluations that address both implementation and effectiveness." By the way, this is a great example of a very high-level or broad policy that will be universally useful. I can't imagine in a context like, say, the federal government, that a simple policy like this would be anything but helpful. If we could encourage Congress, for example, to adopt this simple high-level policy and to delegate to the Executive Branch agencies the responsibility for defining what it means and how it should be operationalized, I think that we would advance the cause of evaluation considerably in our society. Here's another example: "The organization will develop sufficient organization-wide capacity to support evaluation activities." And another: "Evaluations will be designed with input and consultation of key program stakeholders." Notice that some of our evaluation policies may have to do with how we set-up evaluation or how we manage the process of developing evaluation: "Staff will be provided sufficient time to accomplish evaluation-related activities." Some of our policies can be about the capacity that is needed to do evaluation well: "Staff will be provided training in the methodology and use of evaluation appropriate to their program roles." Some policies will relate to the quality or efficiency of evaluation: "Evaluations will utilize the highest quality and most cost-efficient approaches and methods appropriate to the development of programs." This last is another example of a broad or general covering policy. These general policies are essential, especially for their public signaling value, but they are not likely by themselves to be sufficient. We are going to need to develop sub-policies to articulate what these general policies mean. Here's another example: "The highest professional standards will be

utilized to assure 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 professional standards.” We evaluators and other professionals continually debate what the highest professional standards are, and they are likely to evolve over time. But from a policy point of view this general covering policy plays a critically important role and signals that organizations and units that come under the purview of this policy statement are responsible for defending the way they operationalize “highest professional standards.”

## **What Is Evaluation Policy?**

**Who is involved in policymaking?  
Who controls it?**



*“It looks like R.B. is about to make policy.”*

9



How does evaluation policy get made? And, perhaps most important, who is involved in evaluation policymaking? The caption in the cartoon (Slide 9) says, “It looks like R.B. is about to make policy.” Well, I’m going to argue against the R.B. policymaking approach and against the idea of a guy standing alone making policy. In fact, I think that if we use that approach, generally the policy will fail. I’m going to be encouraging us to think about processes and collaborative methods for developing policy as a way of developing sensible policy. We evaluators know how to do this. We’ve been doing this kind of participatory process in our evaluations for some time now and we should adopt such approaches for the development of evaluation policy.

## Why are Evaluation Policies Important?

---

- Evaluation policy has important properties
  - ...a signal → organizing device
  - ...a public stance → transparency
  - ...a mechanism for learning → evaluation policy analysis
  - ...a connection between theory and practice
- Changing evaluation policy is more efficient than changing practices everywhere
  - a systemic approach
- Many evaluation controversies are policy struggles

10



Why is evaluation policy important? I have just a couple of thoughts about this (Slide 10). I think evaluation policy has an incredibly important signaling value. An evaluation policy is a communication mechanism. Think about the metaphor of a lighthouse, a stationary structure that shoots a beam of light 360 degrees and communicates, pretty much on a constant basis, in all directions. A lighthouse is a much better communication mechanism than having people in dinghies trying to row out to the boats that are coming in to warn them about the rocks on the shore. Like a lighthouse, a policy is in one place. It has tremendous signaling power. It is a very efficient way to do things. Another thing about evaluation policy, when it is written policy is that it is a public stance that an organization takes, and in that sense it is a type of transparency, it encourages democracy, it encourages some of the values that I think are really important and that we hold dear. Evaluation policy is also potentially a mechanism for larger learning about evaluation. In Hallie Preskill's AEA Presidential address last year she emphasized the value of viewing evaluation as a type of learning. I think evaluation policy could be one of the key ways for us to learn about evaluation. 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 cumulatively of what kind of policies we think work. Wouldn't it be nice if you were a manager or a policymaker, if you could go to some set of policies – even multiple sets of policies – as a starting point, and say “OK, let me look at how they did it elsewhere.” So, having written evaluation policies enhances our ability to learn about evaluation. Policy also can play a key role in connecting theory and practice in evaluation.

Evaluation policy is potentially an efficient mechanism for changing practice. For instance, if you want to change an evaluation practice in a large organization or in the government, you can do this by going to each local context and making the change locally. But if you are operating in an evaluation policy framework, the easier and more efficient way would be to change the policy once and have that change cascade to all relevant practice contexts. So, potentially evaluation policy can be more efficient.

Another reason evaluation policy is so important is that many, many of our controversies in evaluation today are essentially about policy. I don't want to spend too much time there, but I will just allude to what Lois-ellen Datta said yesterday in her plenary address. She was talking about the U.S. Office of Management and Budget (OMB) and their Program Assessment Rating Tool (PART) system. She described in PART the emphasis that they place on randomized controlled trials as a preferred methodology for effectiveness evaluation. This is essentially an evaluation policy, one that has engendered considerable controversy. 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. Some of you will remember that that was a very difficult conflict for us and consumed considerable energy and angst at one of our conferences. These kinds of issues are fundamentally issues about what evaluation policy should be. So, we can't ignore evaluation policy. It keeps coming back to us.

### Power and Control in Evaluation Policymaking

## **How Can Evaluation Policy be Made?**

### **Authority to make policy versus responsibility for implementing it**



*"Let's switch. I'll make the policy, you implement it, and he'll explain it."*

11



The issue of who controls or has power in evaluation policymaking is a critical one, and helps introduce questions about the roles that various stakeholders play in the process (Slide 11). This cartoon nicely illustrates the multiple roles involved with the caption "Let's switch, I'll make the policy, you implement it, and he'll explain it." When I say to you, "I'll make the policy, you implement it, and you communicate it," I'm certainly saying something about who controls what. But, it also says something about the distinction between the authority to make policy versus the responsibility of implementing it, or the responsibility of communicating about it. These are issues that are going to be challenging to us as we do evaluation policy.

## How Can Evaluation Policy Be Made?

---

### Some Guiding Questions

- Who should be involved?
- How can they be involved most effectively?
- Where is the power and control?
- How do we balance between policy variations and policy “monoculture”?
- What is specified and what is delegated?

12



Here are a few questions that you might think about as we go through evaluation policy development (Slide 12). Certainly, one of the major questions is “Who should be involved?” “How can they be involved most effectively?” I really look forward to watching our field take the many, many years of experience we have with participatory collaborative methods and use this experience not just for evaluation, but for the development of evaluation policy. “Where is the power and control?” is absolutely central to the discussion of evaluation policy. How do we deal with that, how do we address it? “How do we balance between the need for policy variations and pressures pushing us toward policy ‘monoculture’?” By ‘policy monoculture’ I mean by the tendency in some systems for us to establish one set of policies that are applied everywhere. For instance, there is tremendous pressure to have a federal evaluation policy that applies to all federal programs. It feels rational and efficient. On the other hand, if we don’t have variations in policies, we have very little way of determining what kinds of policies work better than other kinds of policies. We don’t have the ability to compare and contrast alternative policies. I’m going to argue that we need to find a balance between our desire for a consistent system and the need for us to learn by examining alternative policies.

One of the key evaluation policy development issues you are going to see in a second is that I believe that a challenge for us in any kind of policy context is to what degree do we state a policy specifically? To what degree do we state evaluation policy at a general level and delegate the responsibility to somebody else to detail the specifics? This issue of delegation is going to be critically important.

## How Can Evaluation Policy Be Made?

**Those with authority may fear sharing it**



*"May I remind you, Jensen, that teacher evaluation by the group has yet to be approved."*

13



Another important issue is illustrated in the cartoon (Slide 13) that shows a teacher saying, "May I remind you, Jensen, that teacher evaluation by the group has yet to be approved." Poor Jensen, obviously just asked a question like, "Well, when do we get to evaluate you?" Jensen, clearly, is not a person in power in this situation. I think one of the things we have to continually remind ourselves of in evaluation policy is that those who have authority may fear sharing it, and that that's not something we can avoid or should avoid as we are talking about how to make evaluation policy.

## 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

14



One of my major intentions in this talk was to describe a potential taxonomy and methodology for evaluation policy. I'm convinced that one of the reasons we have not had a coherent tradition of evaluation policies is that we lack these key components. I have to confess that I did not accomplish one thing I had hoped to do in advance of this address. I had hoped – and this will be no surprise to those of you who know me well – to engage AEA members in a web based concept mapping project designed to develop an initial taxonomy of evaluation policy categories. That is, I wanted us to collaborate on defining the types of evaluation policies we might need in a comprehensive set of such policies. That clearly fell by the wayside. However, two of my colleagues at Cornell, Monica Hargraves and Margaret Johnson, did accomplish a preliminary pilot concept mapping with Cornell Cooperative Extension, about evaluation policies that they are going to be presenting at this conference. But I didn't get my broader project with you accomplished, so what I will show you here is a taxonomy that I pretty much fabricated intuitively. This is my first cut at it (Slide 14). One of the first things that has to happen – and I am sure knowing this crowd, it won't take you long to engage in this – is that you need to criticize my initial taxonomy and revise it. Maybe we will do concept map at some point and try to figure this out in a more structured way (subsequent to this speech, my graduate student Margaret Johnson undertook just such a project as her dissertation work). But I wanted a tentative evaluation policy classification system so I could use it in this talk to give you a sense of what such a taxonomy might look like and how it might be useful. I divided all evaluation policy into eight categories.<sup>1</sup> In this proposed taxonomy, I am basically saying that in a

---

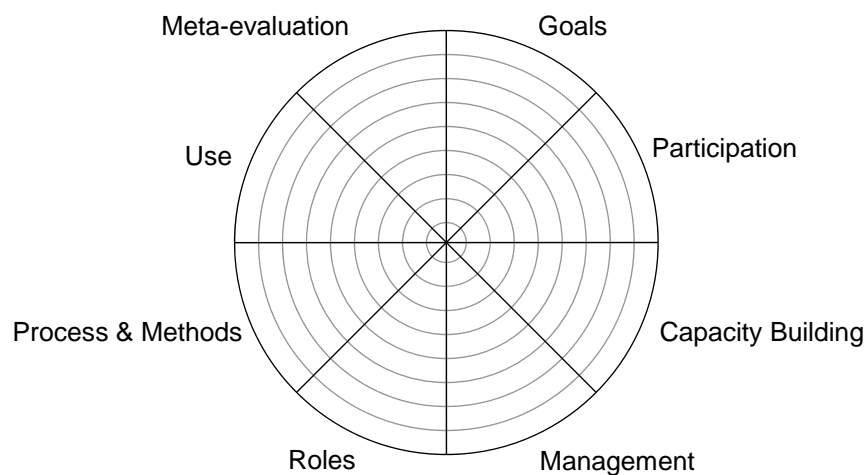
<sup>1</sup> Why eight? Why not seven? Why not five? I'll confess now that the problem for me was entirely driven by my graphic ability with PowerPoint. I think I originally only had seven policy areas, and then I tried drawing a figure with seven pieces of a pie that looked sensible and I just totally gave up. I settled on eight policies essentially because it's easier to divide a circle into eight pieces in PowerPoint rather than seven!



comprehensive set of evaluation policies there would be policies about evaluation goals, policies about evaluation participation, about capacity building, and so on.

Now, in order for you to understand why I wanted this taxonomy and how I intend to use it, I have to tell you a little story. During the past year as a member of Board of AEA we have been working on improving our board processes and aligning our board work more sensibly with the strategic vision of AEA. Little did I imagine last January when we began this work that I would actually learn things from that effort which would make a significant contribution to my understanding of evaluation policy and to this speech. I thought that we would do some nice work on the Board, but that it would not be related to the theme for this conference. In the process of doing our Board work, we began looking at what a Board should do, and we quickly determined that one major role of the Board is to set policy. It seems obvious to the board now, but when we started on this line it was not as clear to us that this policy-making role is central to what a board is. We began looking at the best approaches in the board management literature for dealing with policies on a Board. Anybody who has been involved with the board management field before will know the name John Carver, and will know about his Policy Governance approach to board management. I'm not going to talk about John Carver today, but in Carver's model he has a very interesting graphic device that he uses to convey some important concepts about policy. My intent is to steal that for our discussion here about evaluation policy. My intent is to significantly extend some of Carver's thinking and make it a little bit more appropriate to the evaluation policy context. Carver's focus is on policies in the context of a Board of Directors. I want to talk about policies with respect to evaluation. I think his method, when revised significantly, has some potentially interesting value for us.

## The Policy Wheel

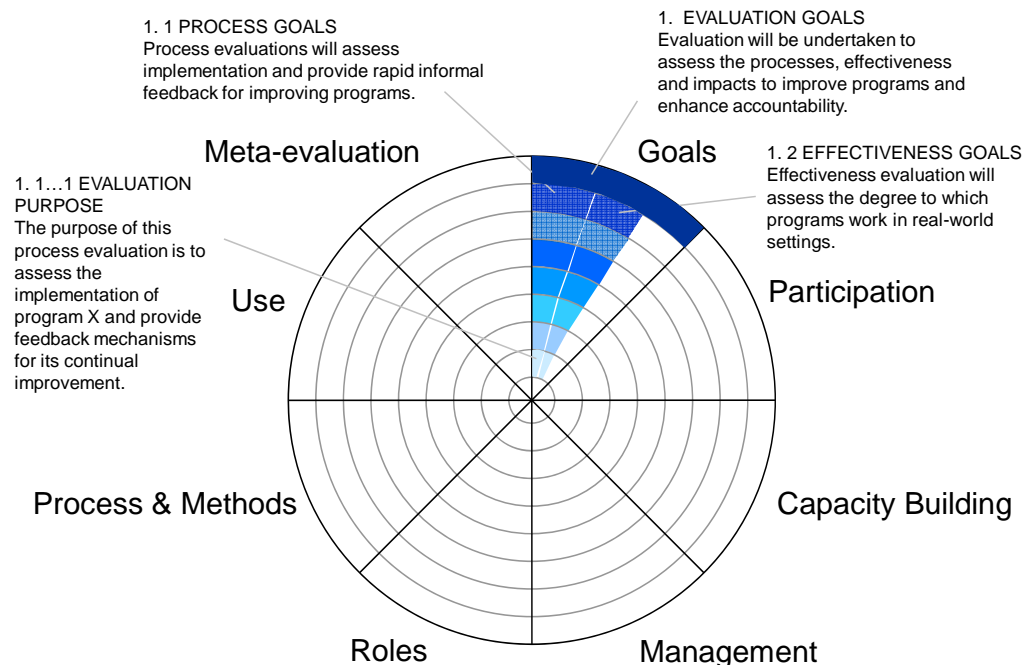


15



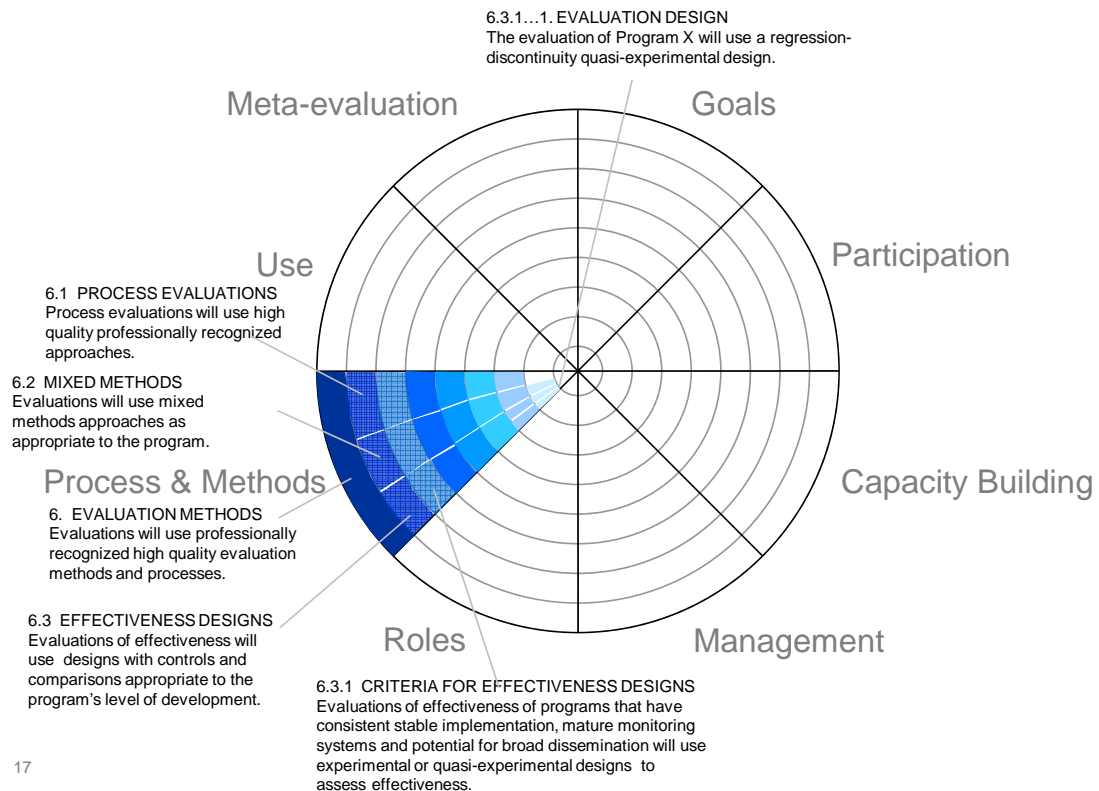
So, here is the method I would propose, adapted considerably from Carver's model. The idea is that you can describe evaluation policy within a simple graphic, a wheel or circle in this case (Slide 15). All of your policies can be located at some place in the wheel. There are different layers of the wheel, from the outermost layer to the center of the circle, suggesting different levels of generality of policy. The most general policy is on the outside, more detailed

sub-policies and sub-sub-policies as you move to more internal layers, and when you get towards the center of the graph, you are at that transition point where policy blends into practice, or procedure, or operationalization. So, the center of the circle is practice, and the outside is the highest-level policy. Presumably this suggests that policy always encompasses practice. On the wheel you can see that I have divided it into eight wedges that correspond to my proposed eight types of evaluation policies. Let's begin by focusing on goals policies and consider a quick example. Again, please keep in mind that I made these policies up as examples and, while they may make sense in a real-world context, we would want to be careful to engage in a process of developing policies and not just adopt these without considerable thought.



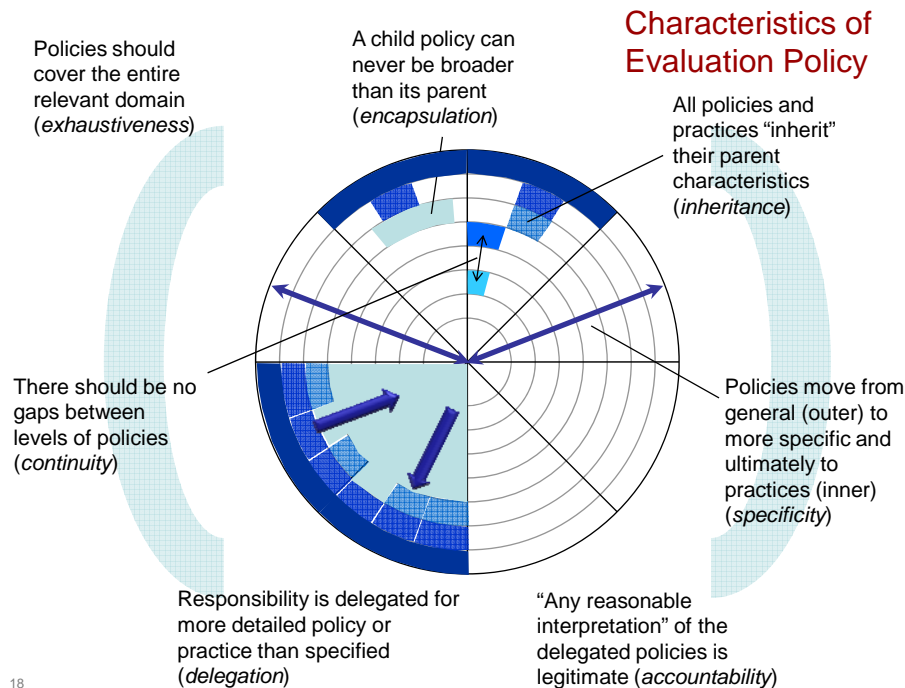
16

What I'm showing in this goals policy wedge is a hierarchy of potential evaluation policies in that area (Slide 16). We begin in the outer layer with the most general goal policy: "Evaluation will be undertaken to assess the processes, effectiveness, and impacts to improve programs and enhance accountability." This is a very broad policy statement. Now, how might we then detail this statement and 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 detailing sub-policies, that is, defining what we mean more specifically by each policy, eventually we are going to get to something that essentially is a description of what we will do in practice. In this example, we might 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 continual improvement."

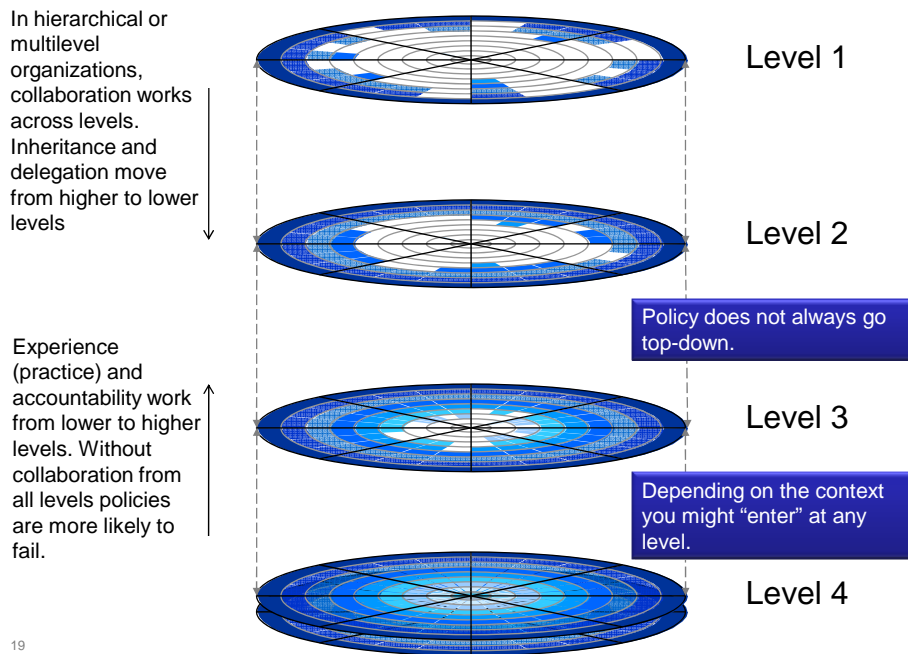


Here's another set of examples, this time in the area of Process & Methods policies (Slide 17). It begins with a general policy requiring "high-quality" evaluation methods and policies, differentiates several different types of evaluations and provides more specific details that describe what constitutes "high quality" methods and processes for these different evaluation types.

I want to argue that this policy wheel provides a useful framework within which we can think about and detail policy.



Let's consider some of the major characteristics of evaluation policy within this framework (Slide 18). I don't have a lot of time to go through all of these, but instead want to convey here some of the important principles that characterize the policy-practice domain. I have already mentioned to you the idea that policies change in specificity as we move from the outer to the inner levels. This suggests that there is a continuum from policy to practice and that the two essentially have differences in degree of specificity rather than differences of kind, what I refer to as the principle of *specificity*. The notion that sub-policies and then practices inherit their outer-level parent characteristics is a principle or characteristic that I call *inheritance*. The idea that this is hierarchical, that is, that broader policies contain within them sub-policies, which contain within them further sub-policies, is what I would term the principle of *encapsulation*. The idea that policies should cover the entire relevant domain and not omit anything important is what I call the principle of *exhaustiveness*. The idea that there should be some continuousness of specification from general policy to specific practice is what I would call the principle of *continuity*. For instance, imagine an agency that has only one very high-level evaluation policy like "We should have regular evaluations of all major programs" and one specific policy (almost at the level of procedure) like "We should evaluate with randomized experiments on every other Tuesday." In this hypothetical case we essentially jumped from a very high-level policy into something that is very, very specific. This is the evaluation policy equivalent of lobbing a grenade into the center of the agency. It is a form micro-managing evaluation practice. And a lot of the problems that we are having in the real world in relation to evaluation policy are related to this type of jump, to the lack of continuity. The principle of *delegation* is the idea that in complex organizations you specify high-level policies at the highest level of the organization and then delegate to subunits who inherit those policies the responsibility for determining more detailed policies. This may involve multiple levels, especially in large organizations, and that delegation carries with it a responsibility for subunits to report back about how they interpreted the high-level policies in making their more detailed sub-policies.



Let's see how this might play out in a real-world context, in this case, the U.S. federal government (Slide 19). At the highest level in the organizational hierarchy – in the case of the U.S. government, this would be the Congress – they might specify some very high-level evaluation policies. They are not going to specify detailed policies. They shouldn't at that level. If they did they would be micromanaging and would likely run the risk of engendering policies that would be inappropriate several levels down and would be actively resisted. As you move down and you work across levels in an organization – this might be in the Office of the President, such as at the Office of Management and Budget – they add details to the policies inherited from above. As you go further down, more detail is added into the policy wheel. As you get towards the center, you get down to the actual operational level. The “policies” at this level virtually turn into short descriptions of the procedures that will be followed in conducting evaluations<sup>2</sup> and the evaluation work is directly shaped by the policies that are inherited. I've got lots more to say here that I haven't got time to say. Certainly, one of the things that I want to emphasize is that this whole multi-level process has to be bidirectional. It should not, it cannot, only go top down. It's got to operate in both directions. Certainly, policy should guide practice. Practice needs to be informed by policy. But the reverse is also true – policy needs to be informed by practice. Good policies have to be put to the test of practice. We need an experiential and empirical practice base to inform policy development.

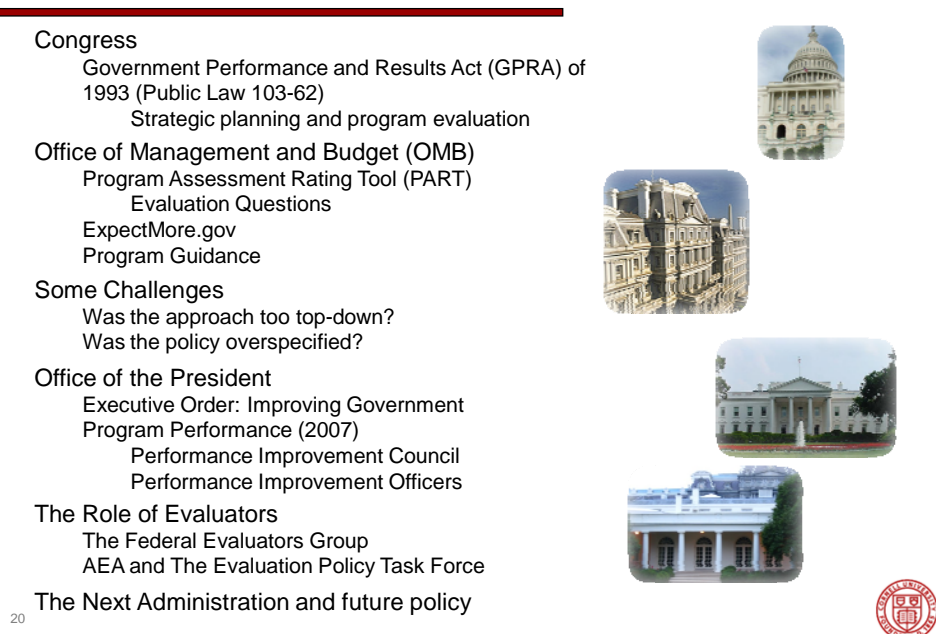
Depending on where you are in the organizational structure you can enter the policy development process at any level of this model. Real life seldom follows a rational model. It is unlikely that we will ever see a policy development effort that begins at the highest level and successively fills in the wheel at lower and lower levels. More likely is that there will be pieces of policies – oftentimes only implicit policies – that already exist at different levels of complex

<sup>2</sup> Notice the relationship between policy and procedure that is implicit in this model. Most organizations already have such a continuum in mind when they use mechanisms like “policies and procedures manuals.” At some indeterminate point, policies become so specific and narrow in scope that they essentially morph into procedures.

organizations. Practicing evaluators are likely to enter such systems several levels below the top of the hierarchy. As we enter and survey the existing policy landscape we'll see that we already inherit a number of implicit or explicit policies. We can then build additional policies as needed at the organizational levels we're operating at. And, we should recognize that policies we adopt may have implications for lower levels of the hierarchy, either now or in the future. That is, our policies in any evaluation help contribute at least to the implicit normative policies that guide evaluation. They act as "precedents" for what was used in previous work.

Incidentally, there is a real opportunity for evaluation consultants in the model I am suggesting. I can imagine a day when evaluators are hired to help large complex organizations to harmonize their existing evaluation policies across levels of the hierarchy and help develop additional new policies where needed. In doing this work, it would be useful for such consultants to be able to consult archives of existing comprehensive sets of evaluation policies developed elsewhere, and I would like to encourage us to begin developing such archives and sharing them.

## Evaluation Policy Case Study



We don't have a lot of time to consider the case study I wanted to go through (Slide 20). I was going to talk about the GPRA law of 1993, a critical law in the history of evaluation. In the figure, you can see that this was developed by Congress who set the general requirement that federal programs had to be assessed for effectiveness. At the next level down, the Office of Management of Budget in the Office of the President, they took that inherited policy and set some more detailed policies, again in that policy circle kind of way. One of the detailed policies manifested itself in the infamous Program Assessment Rating Tool (PART). Two questions of the twenty-five or so of the PART system are specifically about evaluation. These two questions are not detailed enough for implementation, so they decided to develop guidelines that provide even more detail. Here they ran in to some real challenges. For instance, their policy guidance portrayed randomized experimental designs as the preferred methodology for assessing the effectiveness of programs. While randomized experiments may be preferred when the circumstances are appropriate, it is unlikely that a blanket federal policy that requires them would be appropriate for the many and varied contexts and programs involved. So the question



is: did the OMB take an approach that was too “top-down” in this case? That is, was the policy regarding evaluation design over-specified at the OMB level? OMB wanted, quite appropriately, to have a uniform policy across all agencies under its purview, in this case covering all federal programs. How might they have approached this otherwise? Different agencies need evaluation policies that, while consistent with other agencies, appropriately recognize the unique pressures and contingencies in their contexts. If you think about the policy model being offered here, the most salient principle in this example is the principle of *continuity*. In the PART specification, OMB moved from the general GPRA policy set by Congress to one that was likely over-specified in terms of evaluation design. They might instead have concentrated on setting evaluation policies that filled in the gaps between the congressional policy and specific design requirements. For instance, they might have instituted a uniform evaluation policy that all federal agencies must assess the effectiveness of all of their programs in a manner that was appropriate to the program’s stage of development, the need for evaluation information, and that used the most rigorous approach that was feasible. For some agencies and programs this would lead to sub-policies of using randomized experimental designs. But for others, it might be more appropriate to use other methods. By setting a more general evaluation policy at their level, OMB addresses its desire to have a uniform approach across the entire federal government while at the same time helping assure that they are not promulgating an inappropriate “one-size-fits-all” approach. Instead of OMB examiners assessing whether agencies used randomized experiments, they would assess whether the sub-policies and practices of agencies addressed the policy requirements that OMB set.

This “dynamic of delegation” is at the heart of the evaluation policy model I am suggesting here. The challenge in setting evaluation policies at any given level of a complex hierarchy can essentially be reformulated as a problem of delegation. The higher-level unit is responsible for setting policies to the level of specificity that they feel they require. In doing so, they are responsible for making sure that there is a continuity of policy from the ones they inherit to the most detailed ones they ultimately specify. If they over-specify, they run the risk of micromanaging in a manner that will ultimately be counter-productive. Lower-level units will either not be able to fulfill the policies, will actively resist them, or both. There’s some reason to think that this is what has happened with OMB’s PART. Once the higher-level organization specifies a policy, they also need to recognize that they are effectively delegating responsibility to lower-level units to make a “reasonable interpretation” of that policy. Lower-level agencies need to defend the reasonability of their policies and procedures. In this sense, the evaluation policy model offered here is directly linked to the ideas of delegation and accountability.

Getting the level of specificity right is a critical aspect of evaluation policy. OMB is in a position where they operate over the entire executive branch, and their level of policy is going to be limited by that to some extent, but that does not mean that they cannot set policy, delegate responsibility, and have agencies be responsible for explaining how it is they did meet the policies that are being generally specified. In October 2007 the President issued a special Executive Order creating a Performance Improvement Council with representatives from major federal agencies and Performance Improvement Officers in each agency who are getting together and trying to figure out what to do with the next generation of PART evaluation. I think they are beginning to understand that they may not have gone about specifying PART policy in the best way both generally and for evaluation specifically. Evaluators can play a critical role here. For instance, there is a Federal Evaluators group that includes practicing evaluators in many federal agencies that has worked with OMB over the years to encourage sensible evaluation policies.



And, AEA is actively engaged in this work with OMB through its Evaluation Policy Task Force. I think we have made some very important inroads around this kind of policy formulation area. It will continue to be a critically important endeavor for our field.

### Evaluation Policy Challenges

Let's move on now to considering some of the major challenges that we will need to face in developing the idea of evaluation policy. I think it's easiest and most telling to illustrate these challenges with some apt cartoons.

## Evaluation Policy Challenges

### How do we avoid “overformalizing” policy?



21



We've got to watch out about over-formalizing policy (Slide 21). That's a problem that I always have. Some of my friends would say that my name ought to be on the door of the Department of Obfuscation in this cartoon. I fear that we may actually get too excited about specifying evaluation policies and begin developing long, exhaustive, and undoubtedly boring sets of such policies. Finding the right balance between what policies would be specified explicitly and which would evolve as implicit norms is going to be a major challenge for those of us who work in this area.

## Evaluation Policy Challenges

### How do we avoid senseless policy?



*"I'm sorry, I never see anyone personally. However, you may call me from that phone."*

22



Here's another one we need to worry about (Slide 22): how do we avoid senseless or stupid policies? In this example we have a guy sitting across the desk from a visitor saying, "I'm sorry, I never see anyone personally. However, you may call me from that phone." We've all experienced this kind of nonsense in our daily lives. Bureaucracies are especially prone to promulgating these kinds of "Catch-22" conditions.

## Evaluation Policy Challenges

### How do we avoid negative consequences or "side effects" of policy?



23



Here's a cartoon I really love (Slide 23). It shows firefighters are coming outside after putting out a fire in a burning building and their truck is gone. Why is it gone? Because they had

parked the fire truck in a tow-away zone. So, sometimes policies can have unanticipated negative side-effects that can lead to problems. It is certainly possible that we might tie our own hands in evaluation by encouraging more overt written evaluation policy. Such policies, if not well thought through, could actually constrict and limit evaluation rather than encourage it appropriately.

## **Evaluation Policy Challenges**

### **How do we incorporate our core values?**



*"Of course, honesty is one of the better policies."*

24



Here (Slide 24) is one of the most important challenges that I think we'll face: the challenge of ethics. In this cartoon we see a group of businessmen sitting around a boardroom table saying "Of course, honesty is one of the better policies." This is funny in part because it is an obvious hedge on the well-know adage "honesty *is* the best policy." By saying it is "one of the better policies" they are obviously demoting it in value. The challenge for evaluation policy is to get the values right. This will require a difficult balancing act between using the highest ethical standards while ensuring that policies are not just a meaningless organizational cover story that has no correspondence to actual practice. We cannot lose sight of the importance of having evaluation policy-making informed by our core values.

### **Conclusions**

So, let's move on to a few general conclusions (Slide 25).

## Evaluation Policy: Some Conclusions

---

- Get more serious about evaluation policy
- Emphasize the dynamic relationship between evaluation policy and practice
- Encourage written evaluation policies
- Develop a balance between generality and flexibility in evaluation policy
- Acknowledge and address issues of power, hierarchy, delegation, and incentive
- Encourage collaborative and participatory evaluation policy development approaches
- Encourage archiving and sharing of evaluation policies
- Encourage development of a cumulative evidence base about evaluation policies

25



A central and unsurprising message of this talk is that we have got to get more serious about evaluation policy. We should emphasize the relationships between evaluation policy and practice. We should stress the bi-directional nature of the evaluation policy and practice continuum. We have to encourage more written evaluation policies. We have to develop a balance between the desire to have general policies that cover everything and the need for flexibility and adaptation. We have got to address and acknowledge the issues of power, hierarchy, delegation, and incentive; encourage collaboration and participation; encourage archiving and sharing of policies; and so on.

## Evaluation Policy: Some Possibilities

---

- Descriptive studies of current evaluation policies
- Evaluation policy clearinghouses and archives
- Communication and cross-fertilization about evaluation policies
- Evaluation working groups that develop policy
- Evaluation policy analysis
- Software systems for evaluation policy management
- Consulting on evaluation policy
- Engaging the AEA “network”
- AEA’s educational and advocacy role

26



What are some of the possibilities for this newly invigorated emphasis on evaluation policy (Slide 26)? There are lots of potential practical things we can do to further the work in this area. 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 and assess what policies are driving these. Determine which evaluation practices are based on explicit written policies and which are more implicit. Let's start to develop archives and clearinghouses that store this work and share it for others to see. This would certainly help us to communicate and cross-fertilize efforts to develop such policies. In your evaluation work, if you are not already doing so, I encourage you to think about setting up evaluation working groups that think about evaluation policy questions and formulate policies that are appropriate. More formally, we should be exploring the empirical basis for evaluations through use of evaluation policy analysis that compares how alternative policies work and under what circumstances. I think there's a whole potential cottage industry that is possible for evaluation consultants in connection with evaluation policy. For instance, I'm already thinking about the possibility of developing a web-based software tool based on the hierarchical policy wheel graphic. It would allow complex organizations a platform for 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 what other parts of its organization are doing. Of course, we'll need more than just software. I can imagine evaluation policy consultants who are hired by organizations to facilitate the development and management of evaluation policies. Here's where our extensive knowledge of both the field of evaluation and of participatory and facilitative methods could be put to good use. We should also look for opportunities for AEA to take a leadership role in encouraging the development of evaluation policy. One way would be to

use the new web-based platforms that AEA is about to release to enable members to collaborate on and archive evaluation policy work. In the not-too-distant future we should be able to log on to the AEA website and enter a workspace where a group of evaluators interested in this type of work could have a library or archive of example policies, a blog or list that enables them to interact, and so on. And, the emphasis on evaluation policy is directly related to two major current AEA committees – the Public Affairs Committee and the Evaluation Policy Task Force – and their efforts in this regard should have some terrific synergies with broader member activities.

## Evaluation Policy in the Obama Administration

- Evaluation is essential to good federal management
  - Improve program designs
  - Identify start up issues early
  - Share promising implementation approaches
  - Establish, monitor and refine performance standards
  - Develop appropriate and efficient program data collection processes
  - Improve program effectiveness and efficiency
  - Assess program outcomes and impacts
  - Examine dissemination
- Good evaluation **policy** is key to good evaluation **practice** which is key to informed **public policy**
- AEA and its network of evaluators can help develop and manage evaluation policies



I'd like to conclude with a few thoughts about the incoming Obama administration in the U.S. Federal Government (Slide 27). We are meeting here in Denver just two days after Senator Obama has been elected President of the United States. His administration will face 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.

To President-elect Obama we need to say that program evaluation is essential to addressing these issues. It can help answer new questions about current and emerging problems, stop wasteful spending, increase accountability, support major decisions about program reforms and improvements. We are told that he plans to review all government programs to identify those that work and those that don't and to make programs that we need work better than they do now. We need to encourage the new administration not only to examine government programs, but also to examine the way it evaluates them. We need to encourage 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 start up, ongoing implementation, and reauthorization.

For professional evaluators, the opportunity to provide compelling analysis of and practical suggestions for improving public policies and programs is one of the best ways we can enhance our society.

But we evaluators can also benefit from a thorough assessment of how well we perform. For evaluators, as with public policies and programs, there is always room for improvement. I am convinced that one of the major things we can do to organize our own self-assessments and make our work more public and transparent is to renew our efforts to document and assess our own evaluation policies. In this regard, I hope all of you will join me in this effort.

These are indeed challenging times. We hope government officials will make systematic evaluation, including more public articulation and assessment of evaluation policies, a routine part of how they govern and manage. At the same time, I invite all of you to renew your own commitment to your profession and to the policies that guide our work. I hope this national conference will encourage and enable you to do so.