

Mark, M. M., Cooksy, L. J., & Trochim, W.M.K. (2009). Evaluation policy: An introduction and overview. In W.M.K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice*. *New Directions for Evaluation*, 123, 3–11.

1

Evaluation Policy: An Introduction and Overview

Melvin M. Mark, Leslie J. Cooksy, William M. K. Trochim

Abstract

Evaluation policy is of considerable importance, especially in relation to the limited amount of attention it receives as a general topic in the mainstream evaluation literature. Evaluation policies matter for several reasons, among them that they can profoundly affect evaluation practice, they underlie many recent and current controversies about evaluation, and they may be a lever for change that can have far-reaching effects for practice. This chapter gives an overview of several issues regarding evaluation policy, including defining it, identifying possible facets of evaluation policy, describing how it is established, and outlining the potentially greater role for evaluators in shaping the evaluation policies that influence evaluation practice. © Wiley Periodicals, Inc.

This issue of *New Directions for Evaluation* is based on several related beliefs about evaluation policy. First, evaluation policy is a critical concern for evaluation practice, in part because it shapes evaluation practice, thereby both enabling and constraining the potential contribution evaluation can make. For example, if the legislation that establishes a pilot program specifies a particular type of summative evaluation and no other evaluation activities, the pilot program may not benefit from the more developmental



and formative functions of evaluation. At the extreme, evaluation policy may not only enable some forms of contribution and constrain others; bad evaluation policy can have serious negative consequences. For example, consider two areas of work outside the traditional emphasis of most American Evaluation Association (AEA) evaluators. Critics have argued essentially that flawed evaluation policies at the Food and Drug Administration underlay the controversy over Vioxx a few years ago (U.S. Senate, 2004). In the case of auditing, a function related to evaluation, an argument can similarly be made that bad audit policy was at the root of Arthur Andersen's contribution to the rise and fall of Enron (McLean & Elkind, 2003).

A second underlying belief is that evaluation policy deserves more explicit attention in the formal literature and informal dialogue about evaluation. For example, students taking training in evaluation can complete their coursework with little focus if any on evaluation policy. To be fair, in one sense attention to evaluation policy is ubiquitous, in that most writing about evaluation at least implicitly suggests preferences for evaluation policy. At the same time, there is a paucity of literature explicitly addressing the topic of evaluation policy, broadly construed—its origins, how it is developed, the evidence base for evaluation policies, examples of coherent evaluation policies in use, the consequences of having one evaluation policy rather than another, and so on. As another example of the desirability of expanding explicit attention to evaluation policy, it would be interesting to see the preferences of evaluators of various theoretical persuasions. We might speculate, for instance, that certain theorists would highlight methods in their preferred evaluation policies, while others would place more emphasis for policies guiding stakeholder involvement (cf. Alkin, 2004). To take yet another example, more explicit attention to evaluation policy could enrich and improve some of the discourse in which evaluators engage. In particular, evaluators may fail to notice that debate seeming to be about methods is actually better understood as debate about evaluation policies. In the next chapter, for example, Trochim posits that contemporary debate about randomized trials may involve a failure of evaluation policy development, with high-level policy developers creating excessively specific policies that fail to delegate enough authority for decision making at lower levels in the governance hierarchy.

A third belief is that greater involvement of evaluators in development of evaluation policy, if carried out well, could open the door to significant improvements in future evaluation practice. For example, by informing policy makers about what would constitute good evaluation policy, evaluators and their associations might help in constructing evaluation policies that will facilitate evaluations with greater value. We return to this topic later.

Before addressing these and other points, however, we need to clarify what evaluation policy is and is not, especially because confusion can easily arise. By *evaluation policy* we mean policies such as (but not only) high-level rules embedded in legislation that are used to guide the practice of

evaluation. Thus, when Congress passes a bill to carry out a trial of a program called Early Head Start and in the legislation it mandates a randomized trial, evaluation policy has been set. More generally, in Chapter 2 of this issue Trochim defines evaluation policy as “any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation.” In contrast, most of the time when the terms *evaluation* and *policy* are used in the same sentence, the focus is on how evaluation findings (and sometimes process) affect policy in a given area. For example, much has been made of the effects of the Perry Preschool evaluation and other related work on the movement for universal pre-K in the United States (e.g., Schweinhart et al., 2005). In cases such as this, the emphasis is *not* on what we are calling evaluation policy. Rather, the focus in the Perry Preschool case and related ones is on how specific evaluations may influence policy in areas such as early childhood education. In contrast, evaluation policy, as we are using the term, refers to rules or principles that help set the content, characteristics, and context of *evaluation itself*.

The Definition and Scope of Evaluation Policy

Having differentiated “evaluation policy” from “evaluation and its influence on substantive policy,” we need to quickly acknowledge that further attention is called for in defining evaluation policy.

Formal and Informal Policy. One important definitional issue involves whether evaluation policy includes implicit and informal guidance about evaluation, as well as explicit and formal guidance. On the one hand, most observers are likely to agree that evaluation policy has been set when a high level of governance (e.g., Congress, a state department of education, the board of directors of a foundation) formally and explicitly specifies a broad rule that is intended to govern one or more aspects of evaluation. Examples include Great Society legislation that mandated evaluation of new programs and the Department of Health and Human Services “set-asides” that create a pool of funding for evaluation (Shadish, Cook, & Leviton, 1991), the more recent Department of Education priority that specifies a preference for randomized trials for selected funding streams (Donaldson, Christie, & Mark, 2009), the mandate by which monitoring and evaluation findings are reported to citizens including via the Internet in South Africa (National Treasury, 2007), the multifaceted statement of evaluation policy for international development recently put forward by the United Kingdom department working in this area (Department for International Development, 2009), and some foundations’ decision that the evaluation unit will report directly to the foundation president.

In defining evaluation policy, however, an argument exists for a broader scope that would include more informally held rules, as well as the more formally developed policies. Trochim contends in the next chapter that informal, implicit rules should be included under the umbrella of evaluation

policy. Consider, for example, a foundation or state agency where everyone knows that all evaluations should begin with development of a logic model, but no one can point to a written dictate or say when and how this informal rule came to be. Trochim considers these informal rules to be evaluation policies, even if no one can lay out the “policy making” that led to their existence. An analogy to substantive policy making supports this inclusion of more informal, perhaps even emergent, rules as policy. In the area of foreign policy, an old adage states that policy is “made on the wires,” that is, developed in real time in the cables wired back and forth between an embassy and State Department staffers.

Unlike Trochim, several of the other authors in this issue appear to focus on more explicit evaluation policies. This does not, however, mean that the authors would exclude more implicit and informal policies from the realm of evaluation policy. In any event, we support a broader and more inclusive definition of evaluation policy. In addition, the informal-formal distinction may lead to interesting hypotheses, regardless of how expansive a definition one endorses. In particular, an argument can be made that benefits will result when informal evaluation policies are converted to more explicit ones, *if* the process of developing more informed policies is informed, fair, and reflective of the multiple purposes of evaluation. By contrast, poorly developed, explicit, formal evaluation policies may create more problems than would occur in the presence of informal policies only.

Facets of Evaluation Policy. What aspects of evaluation should and do evaluation policies cover? When evaluators are asked to think about evaluation policy, many of them may tend to think first of policies that guide the methods for carrying out an evaluation. Contemporary debates about the place of randomized controlled trials (RCTs) reflect one example in which the focus of evaluation policy is on methods. However, evaluation policy can specify a range of characteristics related to evaluation. What gets evaluated—all programs, certain programs, clusters of related programs rather than individual programs? Which evaluation methods are to be used (and under what conditions)? What is the process by which contingent, situationally responsive method choices are to be made? How are evaluations funded, contracted, overseen? How are evaluators to be selected? Is there some credential that evaluators should have? What are the structural relations surrounding evaluation (e.g., is there an independent evaluation shop in an agency)? How are various stakeholders to be involved, and when? What practices are to be undertaken for reporting, dissemination, and facilitation of use?

Trochim, in Chapter 2, offers an eight-wedge “evaluation policy wheel” intended to capture the various facets that comprehensive evaluation policy should cover. In Chapter 3, Datta offers a related set of questions that evaluation policy could address. A clear message is that evaluation policy can, and in the ideal should, address a number of considerations, ranging from management to method to participation. Of course, in practice actual

evaluation policies will often be incomplete, addressing some but not all of the potential facets.

How Evaluation Policy Is Set

Perhaps not surprisingly, there is not a single, generalizable answer to the question of how evaluation policy is set. Evaluation policy setting will likely vary: across executive and legislative branches; across federal, state, and local governments; across government, foundation, NGO, and private organizations; across the United States and other countries; across formal and informal policies; and across the specifics of one situation versus another. In addition, just as various theoretical approaches that emphasize different actors and process have been proposed to explain the development of substantive policy (e.g., Bennett & Howlett, 1992), alternative frameworks could be employed to account for creation and revision of evaluation policy. Drawing on these alternative frameworks as well as experience, we offer several suggestions about the development of evaluation policy:

- There are multiple ways for evaluation policy to be set, at different times and places.
- Evaluation policy is sometimes used as a lever to accomplish other ends (e.g., delaying general action by mandating evaluation of a policy program).
- At times evaluation policy is an afterthought, tacked on late to substantive policy.
- Without the involvement of evaluators, those who set evaluation policy may not appreciate the multiple contributions that evaluation can bring.
- Often there will be windows of opportunity (e.g., as legislation is being developed) when it is more feasible to influence evaluation policy.
- In many instances, opportunities may exist to influence some facets of evaluation policy (e.g., institutional arrangements of the sort Chelimsky focuses on in this issue) but not all facets of a comprehensive policy.
- Regardless of the process by which it is developed, evaluation policy affects evaluation practice.
- Efforts by evaluators to contribute to the development of evaluation policy ideally will be carried out in ways that also enhance democratic values and good governance.

The Potential Role of Evaluators in Establishing Evaluation Policy

Evaluators can play a role in setting and revising evaluation capacity, acting as individuals, in various networks, and collectively through their professional associations. Well-positioned evaluators, individually and in networks, have been instrumental in developing, maintaining (sometimes in

the face of opposition), and improving evaluation policies in the organizations in which they work. Examples are available in this issue of *NDE*, especially in the contributions by Chelimsky and Leeuw, although the evaluator's contribution is not always explicitly highlighted. Although the action of individuals and informal networks can be powerful, we focus here on professional association involvement, in part because this has been an area with recent activity and with promising potential for the future.

American Evaluation Association (AEA) Activities

In October 2007, AEA announced formation of the association's Evaluation Policy Task Force (EPTF). Authorized initially for two years, the goal of the EPTF is "to assist AEA in developing an ongoing capability to influence evaluation policies that are critically important to the practice of evaluation" (American Evaluation Association, 2007). Eight members were appointed to the EPTF (Eleanor Chelimsky, Leslie Cooksy, Katherine Dawes, Patrick Grasso, Susan Kistler, Mel Mark, Stephanie Shipman, and William Trochim as chair). George Grob, president of the Center for Public Program Evaluation, was contracted to serve as a consultant. A strategic decision was made to focus the initial EPTF on evaluation policies in the U.S. federal government.

In its relatively short existence, the EPFT has been engaged in notable activities. Of particular interest is its involvement with the U.S. Office of Management and Budget (OMB). A key early step was a meeting of the EPTF chair and consultant with Robert Shea, then the associate director of OMB for administration and government performance, and a key person in the agency's design and use of the Program Assessment Rating Tool (PART), a key evaluative mechanism in recent years. Following that meeting, Shea invited detailed comments from the EPTF on a document titled "What Constitutes Strong Evidence of a Program's Effectiveness?" which had been cited approvingly in OMB's PART guidance. The EPTF comments were well received. More recently, the EPTF prepared an "Evaluation Roadmap for a More Effective Government," which was transmitted to Peter Orszag, the new director of OMB in the Obama administration. The EPTF and its consultant have also been involved in several other activities aimed at contributing to federal policy in specific areas and in building relationships and trust to facilitate future work.

As this volume goes to press, the story of the EPTF continues to evolve. For example, it has recently sent recommendations (on request) to both House and Senate committees responsible for health care reform. In its ongoing work, the EPTF continues to seek to influence evaluation policies at the U.S. federal level. As this early experience suggests, there appears to be reason for at least guarded optimism that collective action by evaluators can contribute to more thoughtful consideration of evaluation policy by those who can set it. Moreover, this may prove to be a case in which practice informs theory in the sense that the activities of the EPTF,

and of other similar efforts, help advance understanding of evaluation policy setting.

One of the key challenges in an association's efforts to affect policies is ensuring that the efforts represent the interest of members, without turning the endeavor into pabulum and platitudes. In the case of AEA's EPTF, several steps were taken to facilitate involvement of AEA members, including a special page on the organization's website (<http://www.eval.org/EPTF.asp>), sessions on the initiative at the 2007 and 2008 annual conferences, an electronic discussion listserv, and periodic announcements in the AEA newsletter. Overall, the feedback received from members was quite encouraging of the task force's activities. In addition to contact with members, the EPTF has reported periodically to the AEA board. The EPTF has also included at least two members who are on the board as president, president-elect, or past-president. This has helped ensure the EPTF did not drift from the intentions of the association's elected leadership. Nevertheless, one of the long-term challenges of an evaluation policy initiative involves managing the tension among (1) offering a timely response when a window of opportunity appears; (2) representing well the values and interests of the association, the broader field, and those who may benefit from evaluation; and (3) facilitating development of better evaluation policy while being neither too general to be useful nor too specific to be respectful of the diversity within evaluation practice.

A Note Regarding Origins

This issue of *NDE* has its direct and immediate origins in the 2008 AEA annual conference. Bill Trochim, the 2008 president, selected the topic of evaluation policy for the conference theme. The chapters in this issue, with the exception of the opening and closing ones, were all based on the plenary addresses and expert lectures presented in the conference's Presidential Strand (which was co-chaired by Leslie Cooksy and Mel Mark). There is, however, a longer and more circuitous set of connections that helped lead to the current issue. Evaluation policy, though often discussed in a broader context, has been of considerable interest to AEA boards and presidents for some time. For example, before, during, and after Mark's term as president in 2006, there was frequent conversation about how AEA should "be at the table" when important decisions were being made by federal agencies and other groups about how evaluation should be done. These conversations translated into such actions as the AEA EPTF and the Public Forums that have become a regular feature of the AEA annual conference. The forums are organized by the Public Affairs Committee (PAC), starting with the first in 2006, which was organized by then-PAC chair Bill Trochim, and addressed PART from multiple perspectives. Between involvement in EPTF and PAC (which Cooksy chaired in 2008), Trochim, Cooksy, and Mark have had the opportunity to learn about evaluation policy from such people as

George Grob, EPTF consultant; the other EPTF members; the PAC members; and the participants in the PAC forums, among others.

Several interrelated points arise from this brief and seemingly solipsistic history. First, interest in the topic of evaluation policy has a longer and broader history than that of one year's conference theme. Indeed, we would like to think it is a topic whose time has come. Second, the issue editors owe a debt of gratitude to many colleagues not listed here, among them several past presidents of AEA as well as fellow members of AEA's board, PAC, and EPTF. Thanks to those and others! Third, the editors of this issue wish to express their gratitude for the privilege of being in positions that allow them to try to move things ahead in terms of evaluators' thinking about and acting on evaluation policy. Finally, this brief history, as well as the stories told in several of the chapters in this issue (especially those by Leeuw, Chelimsky, and Stern), offer some hope that the area of evaluation policy is one in which evaluators, individually and collaboratively, can at times make a difference.

Caveats and Conclusions

We hasten to add several caveats about the presentation of evaluation policy. First, although evaluation policy is a topic that applies at various levels of government, and to nongovernmental organizations and various sectors, the chapters in this issue of *NDE* focus on national (U.S. federal) and to some extent supranational (e.g., EEU) units. This is not to imply that evaluation policy is more important, or easier (or harder), at these levels than at others. Rather, given the potentially wide range of territory that could be covered, we chose to focus our efforts (similar consideration has led the EPTF to focus on U.S. federal evaluation policy). Expanding on this focus is one potentially valuable direction for future work on evaluation policy. This may be especially important because organizations may vary tremendously in their capacity and willingness to develop evaluation policies, at least of the more formal and explicit variety. In addition, the best processes for influencing evaluation policy may prove to differ across type of organization.

Second, we have likewise focused primarily on program evaluation (although some chapters veer into policy evaluation and audit). Again, this choice should not be taken as implying that evaluation policies apply to program evaluation and not to other types of evaluation (or to related endeavors). Our sense is that much of what is said here in the context of program evaluation will generalize to other related endeavors, and we encourage exploration or refutation of this suggestion.

A third caveat arises from an earlier observation that there has been a paucity of past work done explicitly on the topic of evaluation policy. In part as a consequence, we do not see the work captured in this issue as an enduring statement. Rather, we hope to stimulate additional work, both conceptual and empirical, that will expand and adjust our current understanding of evaluation policy.

A fourth caveat applies to taking action intended to help shape development of better evaluation policy. It would be naïve to expect that most organizations will, overnight, acquire the capacity and motivation to develop sound and comprehensive evaluation policies. Nevertheless, there is reason for optimism that thoughtful, concerted effort on the part of evaluators can contribute to worthwhile improvement in the evaluation policies that affect our work and the contributions it can make.

References

- Alkin, M. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.
- American Evaluation Association. (2007, October 17). *EPTF announcement letter*. Retrieved April 2, 2009, from <http://www.eval.org/EPTF.asp>
- Bennett, C. J., & Howlett, M. (1992). The lessons of learning: Reconciling theories of policy learning and policy change. *Policy Sciences*, 25, 275–294.
- Department for International Development. (2009, June). *Building the evidence to reduce poverty: The UK's policy on evaluation for international development*. Retrieved July 7, 2009, from <http://www.eval.org/EPTF.asp>
- Donaldson, S., Christie, T. C., & Mark, M. M. (2009). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.
- McLean, B., & Elkind, P. (2003). *Smartest guys in the room: The amazing rise and scandalous fall of Enron*. New York: Penguin.
- National Treasury. (2007). *Framework for managing programme performance information*. Pretoria: National Treasury, Republic of South Africa. Retrieved July 8, 2009, from <http://www.thepresidency.gov.za/main.asp?include=learning/reference/framework/index.html>
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime effects: The HighScope Perry Preschool study through age 40. *Monographs of the HighScope Educational Research Foundation*, 14.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage.
- U.S. Senate (2004, Nov. 18). FDA, Merck and Vioxx: *Putting patient safety first?* Retrieved January 13, 2005, from <http://finance.senate.gov/sitepages/hearing111804.htm>

MELVIN M. MARK is professor and head of psychology at the Pennsylvania State University. He has served as president of the American Evaluation Association and as editor of the American Journal of Evaluation (now editor emeritus).

LESLIE J. COOKSY, incoming president of the American Evaluation Association, is an associate professor at the University of Delaware, where she directs a graduate program in evaluation.

WILLIAM M. K. TROCHIM is professor of policy analysis and management at Cornell University and is 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.