WILEY

ORIGINAL ARTICLE

The future of evaluation policy

Melvin M. Mark[1] | **Nicholas R. Hart**[2]

[1]Department of Psychology, Penn State, University Park, Pennsylvania, USA

[2]The Data Foundation, Washington, DC, USA

Correspondence
Melvin M. Mark, Psychology, Penn State, University Park, PA, USA.
Email: m5m@psu.edu

Abstract
We highlight some key issues regarding evaluation policy, including themes that emerged across chapters of this volume. These topics include what an evaluation policy is, the kind of content that evaluation policies can have, learning agendas (which are an increasingly common component of evaluation policies, especially at the U.S. federal level), the processes by which evaluation policies are developed and implemented, the role of relationships in evaluation policies, and the consequences of evaluation policy. We briefly highlight how the chapters in this volume offer guidance to those involved with developing, implementing, or revising an evaluation policy, especially—but not only—in the U.S. federal context in the wake of legislation signed into law in 2019. Looking to the future, we also share suggestions for further advances with respect to advocacy, accountability, research, and practice related to evaluation policies.

INTRODUCTION

The topic of evaluation policy has received less attention among evaluators than it deserves over the decades since "[m]odern social program evaluation emerged in the 1960s" (Shadish et al., 1991, p.22). Most articles and books about evaluation focus on evaluation methods, evaluation theories or approaches, and even the paradigms that undergird evaluation approaches—far more so than on evaluation policy. The relative inattention to evaluation policy is unfortunate and perhaps surprising, given that evaluation policy can determine whether, when, and how systematic evaluation is done, thereby "enabling and constraining the potential contributions evaluation can make" (Mark et al., 2009, p. 3).

In this chapter, we use the previous contributions to this issue of *New Directions for Evaluation* (*NDE*) as a starting point for identifying a set of key issues and themes regarding evaluation policy. Our comments, and indeed the current issue itself, update an earlier issue of *NDE* (Trochim et al., 2009) and draw on a growing body of recent writing on evaluation policy, much of it reporting empirical research (e.g., Al Hudib, 2018; Al Hudib

- -

& Cousins, 2021; Christie & Fierro, 2012; Christie & Lemire, 2019; Dillman & Christie, 2017; Fierro, 2019; Kinarsky & Christie, 2021; Newcomer & Hart, 2022). Much has changed since the 2009 issue, beyond the recent scholarship on evaluation policy (scholarship that, notably, has come from contributors to this issue of *NDE*). These changes include more widespread development of explicit evaluation policies in agencies and organizations; important legislation at the U.S. federal level, particularly the Foundations for Evidence-Based Policymaking Act of 2018 (the Evidence Act), which was signed into law in early 2019; and ongoing changes in practices related to and emanating from evaluation policies, including those mandated by the Evidence Act.

KEY ISSUES AND THEMES REGARDING EVALUATION POLICY

What is an evaluation policy?

For those not already familiar with the idea of evaluation policy, their first question might be what exactly these are. An answer to this question might refer to the *definition* of evaluation policy and the kind of *content* typically included in an evaluation policy. Chapters in the issue of *NDE* help clarify both.

A commonly used definition was presented in a chapter in the 2009 *NDE* issue on evaluation policy. Drawing on his 2008 Presidential Address at the American Evaluation Association (AEA) annual meeting, Bill Trochim defined evaluation policy as "any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation" (p. 16, italics from the original removed). Al Hudib and Cousins (this volume) revised Trochim's definition based on their review of 52 evaluation policies from international aid organizations, largely by expanding beyond Trochim's focus on "doing evaluation." They define evaluation policy as "any principal, norm, standard, framework, strategy or rule that is used to guide an organization's decisions and actions when planning, designing, doing, reporting or using evaluation and that is directed to develop or enhance organizational capacity." A simplified version, omitting the emphasis on organizational capacity, might define an evaluation policy as consisting of the principles, guidelines, or other dictates that guide an organization's and its members' understandings and actions regarding the planning, conduct, or use of evaluation.

Beyond the definition of evaluation policy, those unfamiliar with the concept might wonder what goes into an evaluation policy. Expanding on the eight categories that Trochim (2009) tentatively identified, Al Hudib and Cousins (this volume) identified 10 categories of content in evaluation policies. These are: the goal of evaluation; the definition of evaluation; roles related to evaluation; management; evaluation norms, standards, criteria, and principles; planning and processes; evaluation use; meta-evaluation; evaluation capacity building (ECB); and competencies and ethics.

Drawing on the sample of evaluation policies they examined, Al Hudib and Cousins also provide examples of the specific content corresponding to each category. These examples illustrate the kind of specific guidance that an evaluation policy can provide. For instance, with respect to evaluation goals, Al Hudib and Cousins report that the majority of the evaluation policies they examined "identified the promotion of accountability as *the* principal goal." Learning was specified as a goal for evaluation in almost half of the policies (and in all of the more recently developed policies). Several of the policies also identified the promotion of an evaluative culture as a goal. Most other chapters describe aspects of the content of evaluation policies, with Irwin and Nightingale providing a table including examples from four U.S. federal agencies.

Foreshadowing a point made later, one could imagine future research that examines whether differential consequences arise depending upon which goal(s) an evaluation policy emphasizes. How, if at all, do organizations whose evaluation policies focus on accountability differ from those who focus on learning, in terms of how evaluation is conducted, or used, or how well evaluation capacity is built? How do the policies vary by organizational structures, resource availability, and even nationality? Given the pioneering work of Trochim (2009) and the more recent work of Al Hudib and Cousins (2021; this volume), attention to evaluation policy in the future might focus less on general categories and definitions. Instead, more attention could be given to a policy's specific content and how best to tailor policies to particular organizational needs and missions.

Learning agendas

Learning agendas are a noteworthy example of specific content within a broader evaluation policy. The chapter in this issue by Newcomer, Olejniczak, and Hart focuses on learning agendas. Irwin and Nightingale (this issue) also address learning agendas, detailing the creation and use of one in the Department of Labor. Both chapters also refer to other U.S. agencies that had learning agendas in place before the Evidence Act. That act has been stimulating considerable attention in the federal sector to these strategic plans for evaluation. We expect to see increasing use of learning agendas and other sectors, including state agencies and foundations.

An interesting implication arises from the increasing prominence of learning agendas as part of an overarching evaluation policy. That is, learning agendas shift the focus of the policy's attention. For earlier evaluation policies without a learning agenda requirement, the policy's guidance focused primarily on the nature and conduct of individual evaluations. Learning agendas, in contrast, elevate the focus of the evaluation policy, in part, to a portfolio of evaluations and other sources of evidence that address high-priority questions for the organization.

We see this as a valuable shift. Not only does this shift in focus place evaluation in a broader context of evidence used in service of learning and decision making. In addition, as suggested by Al Hudib and Cousins' (this volume) discussion of linkages between evaluation policy and ECB, learning agendas should provide leverage for increasing evaluation capacity. This may prove especially true for increasing the capacity to *use* evaluation, as Newcomer and colleagues suggest. In part, this could occur because the process of developing a learning agenda itself may increase the likelihood that potential evaluation users will use evaluation findings.

There are two clear opportunities for maximizing the alignment of the learning agenda to other activities in organizations. First, calibration with an organizational strategic plan is one way to ensure the items in the learning agenda sync with overarching goals and priorities influenced by political or senior-level leaders. Indeed, establishing the political or policy goals and priorities may be desirable *before* formulating the learning agenda questions and approaches. Second, in many parts of government a particular policy matter or issue may span the domain of multiple agencies or organizational units. For example, the topic of homelessness is addressed in the federal government by multiple agencies. Organizing learning agendas across major themes through coherent interagency processes and topics could help ensure the alignment of resources for evaluation and evidence building, while aligning core capacity and infrastructure to solve the priority informational needs for decision-making.

The process of developing and implementing an evaluation policy

Contributors to this issue collectively give considerable attention to the development and implementation of an evaluation policy. This is the primary focus of the chapter by Fierro and her colleagues. The development and rollout of evaluation policy is also illustrated in the case study by Irwin and Nightingale. At the same time, Epstein and colleagues address it in terms of the agency workforce and various pragmatics. In some chapters, particularly Newcomer et al., attention is given to the development and rollout of part of an evaluation policy, specifically, a learning agenda.

An important observation, by Fierro et al., Irwin and Nightingale, and other contributors, is that the process of developing and implementing an evaluation policy may have important consequences beyond the existence and deployment of the policy. Akin to process use (Patton, 2008), the engagement of various parties, including agency leadership, in policy development and rollout may enhance their understanding of and commitment to evaluation, which in turn may increase the likelihood of evaluation use (Newcomer et al.; Irwin and Nightingale, this issue). Beyond the components and specific content of an evaluation policy, the way the policy was developed and implemented may determine its effects on evaluation capacity.

Another important point throughout this issue is that, for an evaluation policy or learning agenda to have positive effects, it cannot be treated as a checklist, compliance-type activity. Instead, it needs to be undertaken authentically to serve the agency's overarching mission and intended beneficiaries. A meaningful involvement of managers and agency leadership is important.

Evaluation policy in context, and as implemented

Most contributors to this issue of *NDE* pointed out that one size does not fit all when it comes to evaluation policy. Context matters. Evaluation policy development, content, implementation, application, and revision may all be done differently depending on the unit's mission, size, history, and other considerations. With the increasing involvement of federal agencies in evaluation policy in the wake of the Evidence Act, more specific and more evidence-based guidance might become available for subsequent efforts to develop, employ, and revise evaluation policies in a context-sensitive fashion.

This wish is more likely to come true if future investigators attend not only to the espoused policy (as written) but also to the policy in use. For example, Epstein and colleagues point out that policy may call for oversight of evaluation contractors by agency staff, but the degree and intensity of oversight might vary. Evaluation policy in use would probably be a fruitful topic for future research on evaluation.

Future attention might also be given to how evaluation policy intersects with and is influenced by other policies and regulations. In discussing the role of federal staff in carrying out evaluation policy, Epstein and colleagues point to the constraints created by rules regarding contracting, clearances for data collection, and security requirements. These are part of the context in which federal agencies' evaluation policies are implemented. Even if such constraints are in a sense consistent across agencies, the way agencies deal with them (e.g., whether or not they have specialized staff to implement the Paperwork Reduction Act) may not be. More generally, the infrastructure for evaluation, including the nature and the stability of funding (Epstein et al. and Irwin and Nightingale, both this volume), may be vital factors in how successfully an espoused evaluation policy is translated in practice.

The role of relationships

On a piece of paper or a computer screen, an evaluation policy might seem like an abstract thing. Several chapters highlight the human role, with evaluation policy intertwined in the complicated and sometimes messy world of human relationships. These chapters particularly highlight the role of relationships in the implementation and ultimately the consequences of evaluation policy.

Al Hudib and Cousins note that, in many evaluation policies, the definition of evaluation describes how it differs from related functions such as performance measurement. We see this not as an abstract definitional issue or a simple power-play to stake out ground, but rather as a way of helping clarify the relationship among different functions and actors within the agency. The Irwin and Nightingale chapter nicely illustrates, in the context of the Department of Labor's evaluation policy, how clarifying the role of evaluation relative to other evidence-informed activities, helped get evaluators a seat at tables from which they otherwise may have been excluded. Newcomer and colleagues and Fierro et al. indicate the importance of evaluators developing relationships with and across senior leaders in the agency, which can be enhanced by the collaborative development of an evaluation policy and/or learning agenda. Epstein and associates also suggest that an important role in implementing an evaluation policy is "to lead and coordinate an evidence and evaluation Community of Practice (CoP) for their agency."

FUTURE DIRECTIONS

This section offers suggestions for future efforts involving advocacy, research, and practice related to evaluation policies.

Advocacy

Absent external mandates to have one, the fact that an evaluation policy exists in an organization is likely a testament to the presence of one or more organizational members who successfully advocated for such a policy. Beyond the important successes of these often unsung heroes scattered across various organizations, more formal, collective efforts at advocacy have occurred. In particular, since 2007, AEA has had an Evaluation Policy Task Force (EPTF). As described in Chapter 1 of this issue, the EPTF has focused on U.S. federal level evaluation policy. Chapter 2 of this issue contains the most recent version of one EPTF product, known in short as the Roadmap. An earlier version of the Roadmap was published by AEA in 2009, following an extensive drafting and revision process, which included input from the AEA membership and board, subsequent revision, and review and approval by AEA leadership.

At the U.S. federal level, the next several years are likely to involve advocacy regarding implementation of the evaluation-related features of the Evidence Act. Advocacy will likely also lead to efforts to ensure that funding and personnel are made available for the basic capacity to implement these new evaluation policies. For example, funding requests for through congressional appropriations processes or advocacy in agency and White House budget formulation processes could be productive avenues for ensuring that evaluation officials and capacity are adequately resourced across government agencies.

Another likely area involves encouraging the inclusion of evaluative activities in major federal initiatives. An important potential direction would involve support for high-quality

evaluation policy at the state, tribal, and territorial levels and nonprofit and other sectors. Given the capacity limits of the EPTF, such expansions would probably require partnerships with others in the lead, or activities by others independent of the EPTF. Perhaps this issue of *NDE* will help stimulate activities supporting sound evaluation policies in various domains in which systematic evaluation is conducted.

Accountability

With new mandates and written policies, there is an emerging opportunity to support accountability and oversight of organizations and agencies as they implement evaluation activities. While Congress can provide some oversight of the Evidence Act, an opportunity arises in professional associations such as AEA and other nonpartisan entities. These third parties can observe the implementation of key evaluation policies and help identify challenges, limitations, gaps, and shortcomings; this in turn can promote adequate oversight of the law and consistent implementation of evaluative thinking and the standards proffered by OMB.

Similarly, increased efforts to draw attention to success stories and achievements would provide positive accountability, acknowledging the leading agencies' and officials' implementation of effective organizational policy. AEA could, for example, use its award process to recognize the leadership of evaluation officials in government.

Practice

The chapters in this issue of *NDE* should be of considerable interest to those engaged in developing, implementing, and revising an evaluation policy in their organization. Collectively, this issue can help guide understanding and action regarding the following:

- what an evaluation policy is;
- the general issues or categories an evaluation policy might address;
- the way one might go about developing (or revising) and rolling out an evaluation policy;
- the multiple roles that agency staff can and should play in developing and carrying out an evaluation policy;
- the potential place of a learning agenda requirement in an evaluation policy, including ways to develop a quality learning agenda; and
- many of the nuts and bolts of developing and effectively using an evaluation policy in an actual organization.

We look forward to hearing about the ongoing and upcoming development and implementation of evaluation policies. We hope future reports will include case studies, ideally some with multiple case analyses, sensitive to variations in context and evaluation policy content. Such efforts should help guide the next generation of work on evaluation policies.

Research on evaluation

Several of the chapters in this issue of *NDE* report findings from systematic research on evaluation. We believe that evaluation policy will be a fruitful arena for research on

evaluation for some time. The chapters in this issue should provide a wealth of ideas for future research on evaluation policy.

Some potential research questions are descriptive, such as how federal agencies new to formal evaluation policy go about developing and implementing a policy in response to the evidence act, how evaluation policy in use varies from espoused policy, and how other policies and regulations constrain or enable successful implementation of evaluation policy. Other potential research directions focus on the potential consequences of an evaluation policy (or specific aspects of it). For example, one might attempt to assess whether (and how) the development of an evaluation policy, or the creation of a learning agenda, contributed to evaluation capacity and evaluation use. As another example, future research might compare, say, organizations whose evaluation policies emphasize only accountability as the goal of evaluation versus those organizations whose policies emphasize learning; the research could examine whether the kind of evaluations conducted, and the kinds of evaluation use differed in the two sets of organizations. Again, there is a bounty of potential research questions across this volume.

CONCLUSION

The current issue of *NDE* has reviewed empirical, conceptual, legislative, and practice developments related to evaluation policy. We hope that, in addition to updating readers on evaluation policy, this closing chapter plus the other chapters help point the way to productive future directions.

Although the chapters focus primarily on the U.S. federal government, the volume has important implications for the broader evaluation community. Whether formalized in a document or not, most of the organizations with which evaluators work have some form of evaluation policy, however informal or incomplete it may be. For evaluators who work with organizations without an explicit evaluation policy, chapters in this issue may help inspire and guide them to encourage, and perhaps lead, the development of a more formal policy. Moreover, given that practices in the federal government often influence practices elsewhere, the Evidence Act's mandates for an evaluation policy and a learning agenda may affect organizations in other sectors.

Finally, we hope that the current issue of *NDE* will help increase evaluators' appreciation of evaluation policies and their importance. Evaluation policies matter. Evaluation policies matter for reasons including that they help determine what form evaluation takes, and they may influence how relevant parties relate and react to evaluation. Evaluation policies thereby enable or constrain the contributions that evaluation can make. In his Presidential Address to AEA and a subsequent article, Will Shadish noted that evaluation theory should be part of every evaluator's identity. Shadish (1998, p.1) said, "Evaluation theory is who we are." To this we add: Evaluation policy is what we can do.

ORCID
Melvin M. Mark https://orcid.org/0000-0003-4545-1553
Nicholas R. Hart https://orcid.org/0000-0003-3541-4337

REFERENCES
Al Hudib, H. (2018). *The role of evaluation policy in organizational capacity to do and use evaluation* [Unpublished Ph.D. dissertation]. University of Ottawa, Ottawa.
Al Hudib, H., & Cousins, J. B. (2021). Understanding evaluation policy and organizational capacity for evaluation: An interview study. *American Journal of Evaluation.* https://doi.org/10.1177/10982140211007573

Christie, C. A., & Fierro, L. A. (2012). Evaluation policy to implementation: An evaluation of scientifically based research in practice. *Studies in Educational Evaluation, 38*(2), 65–72.

Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation, 40*(4), 490–508. https://doi.org/10.1177/1098214018824045

Dillman, L. M., & Christie, C. A. (2017). Evaluation policy in a nonprofit foundation: A case study exploration of the Robert Wood Johnson Foundation. *American Journal of Evaluation, 38*(1), 60–79. https://doi.org/10.1177/1098214016642864

Fierro, L. A. (2019). Research on federal evaluation policies. Paper presented at the Annual Meeting of the American Evaluation Association.

Kinarsky, A. R., & Christie, C. A. (2021). Analysis of evaluation policies in the philanthropic sector. *American Journal of Evaluation*. https://doi.org/10.1177/1098214021989435

Mark, M. M., Cooksy, L. C., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. In W. M. K. Trochim, M. M. Mark, & L. C. Cooksy (Eds.), *Evaluation policy and evaluation practice* (pp. 3–11). Jossey Bass.

Newcomer, K., & Hart, N. (2022). *Evidence-building and evaluation in government.* Sage

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

Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation, 19*(1), 1–19

Shadish, W. R., Jr., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Sage.

Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. In W. M. K. Trochim, M. M. Mark, & L. C. Cooksy (Eds.), *Evaluation policy and evaluation practice* (pp. 13–32). Jossey Bass.

Trochim, W. M. K., Mark, M. M., & Cooksy, L. C. (Eds.). (2009). *Evaluation policy and evaluation practice.* Jossey Bass.

AUTHOR BIOGRAPHIES

Melvin M. (Mel) Mark, Professor of Psychology at the Pennsylvania State University, was co-editor, with Bill Trochim and Leslie Cooksy, of the 2009 *New Directions for Evaluation* issue on evaluation policy, collaborated with Trochim on the establishment of the AEA's Evaluation Policy Task Force, and has been a member since its inception, including time as co-chair.

Nicholas R. Hart is President of the Data Foundation and Chair of the American Evaluation Association Evaluation Policy Task Force. He served as the Policy and Research Director of the U.S. Commission on Evidence-Based Policymaking. He previously worked at the White House Office of Management and Budget during the Bush, Obama, and Trump Administrations.