Article

Why Evaluation Theory Should Be Used to Inform Evaluation Policy

Christina A. Christie[1] and Sebastian Thomas Lemire[1]

Abstract

Many organizations now have evaluation policies. Because formal evaluation policies intend to frame evaluation practice, it is important to understand the aims of policies and their influence on practice. Evaluation theory is also intended to guide practice; therefore, the extent to which policies are informed by theory is also important to consider, given the related purposes of evaluation theories and policies. Because evaluation policies have emerged with greater occurrence relatively recently, the conceptual and empirical work on the theory–policy–practice connection is limited. Our aim in this article is to flesh out the intersection between evaluation theory, policy, and practice. As illustrative examples, we discuss the three policies in terms of methods, use, and valuing—three core dimensions of evaluation theory. The article concludes with a set of reflections on the different strategies for and implications of a stronger integration of evaluation theory in evaluation policies.

Keywords

evaluation theory, evaluation policy, evaluation practice, policy analysis

Introduction

Shadish (1998) argued in his American Evaluation Association (AEA) Presidential Address 20 years ago: "Evaluation Theory Is Who We Are." Expanding on this declaration, Shadish noted several reasons why evaluation theory is so central to the field. Most notably, he proposed, evaluation theories are united in that all have evaluation practice as the principal concern. In this way, evaluation theories collectively serve as a unifying body of knowledge around which to discuss and develop our practice. Of equal importance, evaluation theory also addresses the topics and issues that are of great concern to evaluation as a field including issues related to methodology and methods, use and utilization, and valuing. In concluding on these observations and with particular relevance for the present article, Shadish finally remarked on the foundational bond between evaluation theory and practice, stating, "There is no evaluation theory without evaluation practice, and

[1] University of California, Los Angeles, Los Angeles, CA, USA

Corresponding Author:
Christina A. Christie, University of California, Los Angeles, 2022B Moore Hall, Los Angeles, CA 90095, USA.
Email: tina.christie@ucla.edu

there is no profession of evaluation without evaluation theory; as without evaluation theory, evaluation is simply an applied social science methodology" (Shadish, 1998, p. 13). Evaluation theory is who we are, what we preach, and (at least ideally) what we practice.

Yet, as empirical work has shown, evaluation theory does not predictably transfer into practice (e.g., Christie, 2003, among others). The delta between evaluation theory and practice is real. There are likely several reasons for this theory–practice gap. For one, the importance and role of evaluation theory in evaluation is not without contention. Notable evaluation theorists such as Michael Scriven have argued that it is possible to conduct a good evaluation without explicitly using evaluation theory (Scriven, 2004).

Accordingly, and perhaps especially in a practice-oriented field, the role and relevance of evaluation theory is not "a given." For another, evaluation theory often does not provide procedural guidance, that is, step-by-step advice on how to design and conduct evaluation. By their very nature evaluation theories are broadly applicable, and at least somewhat removed from local contingencies. Consequently, evaluation practitioners can be guided by evaluation theory and still left ample room to interpret and translate these theories into local practice. Because evaluation theory does not automatically transfer into practice, we maintain that there has to be a mechanism or driver for an evaluation theory to gain traction in evaluation practice.

One such mechanism, and one that seems to be growing in significance, is evaluation policy. Not until the last decade or so has the discussion of evaluation policies been of interest to the field of evaluation. As with many significant developments in evaluation, evaluation policies emerge from practice, responding to a need for organizations and government agencies to formally guide evaluation practice. In this way, the evaluation policies importantly may be viewed as a reflection of an organization's values related to evaluation, its procedures, and its potential influences. Formal evaluation policies (especially those of major commissioners of evaluation) serve to frame a significant and still growing part of evaluation practice.

The central role of evaluation policies is easily made. As Trochim (2009) eloquently observes

> Evaluation policies profoundly affect the day-to-day work of all evaluators and ultimately the quality of the programs they evaluate. Many recent and current controversies or conflicts in the field of evaluation can be viewed, at least in part, as a struggle around evaluation policy. Because evaluation policies typically apply across multiple evaluations, influencing policies directly may have systemic and far-reaching effects for practice. (p. 14)

Motivated by these observations, the overarching aim of the present article is to argue for a stronger and more explicit integration of evaluation theory and evaluation policy. If evaluation theory is who we are, and if evaluation policies reflect what an organization or entity's evaluation practice is intended to be, then the deliberate and reflective integration of evaluation theory into our evaluation policies becomes central.

The position we hold is that evaluation policies serve as an important connector between evaluation theory and evaluation practice. More specifically, evaluation policies serve as a transformative mechanism for translating and perhaps more importantly situating evaluation theory in the organizational, political, and cultural context of an organization. In order to do this, in order to bridge the theory–practice divide, we argue that the theory integration has to be purposeful and explicit, emphasizing adaptation over adoption.

This type of theory integration holds merits for evaluation theory, policy, and practice. Evaluation theory may strengthen evaluation policies in several distinct ways. For one, and informed by Shadish (1998), evaluation theory may serve to clarify key concepts and relatedly support a shared language around evaluation across diverse organizations and policies. Another benefit of integrating evaluation theory more firmly into evaluation policies is that it promotes a better understanding of

the contingency of evaluation theory. Being explicit about which theories have been infused into policies will provide the evaluation community with a new and different opportunity to advance our thinking on which theories, elements of theories, or theoretical principles are best used under particular conditions. Finally, and emerging from the former two, evaluation practice is likely to be more informed (if not better informed) by evaluation theory—a unifying body of knowledge around which to discuss and develop our practice (Shadish, 1998).

The purpose of the present article is primarily conceptual—setting the scene for and motivating further empirical work on this important topic. Toward these aims, the article directs attention to the evaluation theory and policy connection, an aspect of the theory–policy–practice nexus that as of this writing has received little scholarly attention. Moreover, the article provides clarification on the role of evaluation policies as transformative mechanisms between evaluation theory and evaluation practice. Finally, the article makes the case for purposeful, explicit integration of evaluation theory into evaluation policy, specifying in extension the potential benefits of this type of theory integration for evaluation theory, policy, and practice.

The article is structured according to three sections. In the first section, we consider the meaning of evaluation theory, evaluation policy, and the notion of the theory–policy–practice nexus. In extension of the latter, we propose that how we frame the evaluation theory–policy relationship holds import and in effect demands to be further unfolded and conceptualized. In the second section, we discuss the evaluation policies from three large evaluation commissioners, illustrating similarities and distinctions in how evaluation theory is emphasized and interpreted reflect local organizational, political, and cultural contexts. Finally, in the third section, we return to the overarching aim of the article and make the case for a stronger integration of evaluation theory in evaluation policy.

The Theory–Policy–Practice Nexus

Before advancing our argument, a few words on the conceptual building blocks comprising the theory–policy–practice nexus. The many different meanings and connotations associated with the terms evaluation theory and evaluation policy call for conceptual clarity.

Evaluation Theory

Evaluation theory, referred to by some as evaluation approaches or models, prescribes how and with what purpose to conduct an evaluation. Thus, we have evaluation theories that are intended to promote the utilization of evaluation results (Patton, 1978), encourage stakeholder empowerment (Fetterman, 1994), or bring attention to context–mechanism–outcome configurations (Pawson & Tilley, 1997), to name but a few. These theories are often informed by the theorists' practical experiences with evaluation and are always influenced by their philosophical orientation and theoretical persuasions (Christie & Alkin, 2012). As a collective theoretic knowledge base, evaluation theory necessarily emerges from a system of interdisciplinary cooperation. Thus, while all theorists are theorizing about how to best realize what she or he believes to be the primary purpose of evaluation, these prescriptions are informed by and developed in dialogue with other theorists.

While prescriptive in nature, and as mentioned above, evaluation theories do not provide the same type of procedural guidance as do situation-specific guidelines and principles. One reason for this is that evaluation theories are not fully contingent on local contexts or settings. Rather, evaluation theories are in scope and nature intended for application across a broad—and often unspecified—range of contexts. In this way, evaluation theories are perhaps best viewed as aids in thinking or as providing a working logic to assist evaluators in designing evaluations, selecting procedures and methods, providing a rationale for procedures and methods used, and distinguishing evaluation from other activities such as applied research.

Tracing the main strands of evaluation theories, three dimensions are discernible: methods, use, and the role of valuing and values in evaluation (Christie & Alkin, 2012). The authors depict these three dimensions of evaluation theory in a "theory tree" that is intended to describe the conceptual foundations of evaluation theory as the "roots" of the tree, with each root grounding one of the three dimensions or "branches" of the theory tree (Christie & Alkin, 2012). Because we will use these dimensions to examine three evaluation policies in the next section of the article, we will succinctly describe the three branches of the theory tree.

In the first version of the evaluation theory tree, Alkin and Christie (2004) note: "First, there was methods" (p. 17), highlighting the centrality of the methods dimension in evaluation theory. This middle branch of the theory tree focuses on the methodological "habits of mind" by which evaluation is promoted to be an activity intended to build knowledge. At root, this dimension is grounded in systematic social inquiry, which is concerned with employing the most rigorous (as described by postpositivist philosophers) techniques for examining the impacts of social programs.

In contradistinction, the use dimension of the tree, which provides a central motivation for evaluation as a mechanism for improving programs and society, is rooted in and motivated by social accountability. Accordingly, the use branch that extends from this root is concerned with how evaluation information is used, by whom, and for what purposes. Principles of practical stakeholder engagement—who, why, and to what end stakeholders are included—are central here too (see Cousins, Whitmore, & Shulha, 2013).

The third dimension, valuing, addresses the ways in which value is placed on the evaluand. Christie and Alkin (2012) argue that there are two perspectives that inform the valuing process, one that argues for a subjectivist and the other an objectivist (as understood in the philosophical sense of the terms) approach to valuing. The subjectivist approach reasons that value judgments should be based on "subjective meaningfulness" because all human behavior is governed by subjective factors, while an objectivist approach maintains that value judgments be made based on "publically observable facts" (Diesing, 1966, p. 124). Several important ideas about the importance of the prominence of traditionally underrepresented and marginalized groups in an evaluation have grown from the subjectivist principles described on the valuing branch. On the objectivist side, using information that can be quantified to determine the worth of an evaluand, such as various cost analysis approaches, has also had a great influence on the field.

Evaluation Policy

With this categorization of evaluation theory as our backdrop, let us then turn to evaluation policy. Despite the widespread emergence of evaluation policies, the topic has received minimal scholarly attention. One noteworthy exception is the special issue of *New Directions for Evaluation* focusing on evaluation policy in the United States and in Europe and the role it plays in practice and the field more generally (Trochim, Mark, & Cooksy, 2009). In this volume, Trochim (2009) states that "An evaluation policy is any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation" (p. 16). As Trochim expands, an evaluation policy may serve as a *communication mechanism* signaling to people inside and outside an organization "what evaluations should be done, what resources expended, who is responsible, how they should be accomplished, and so on," that is, what evaluation is or should be within same said organization (p. 17). In this way, Trochim notes, "evaluation policies make evaluation a more transparent and democratic endeavor" (p. 17). An evaluation policy can also function as a *learning mechanism* by providing grounds for "cumulative knowledge about what kinds of policies that appear to work under various circumstances," in effect, allowing us to "learn about the connection between evaluation theory and practice" (Trochim, 2009, p. 18).

Extending on Trochim, we hold that the primary of purpose of an evaluation policy has to be to provide guidance on how, when, and with what purpose evaluations are carried out within an organization, that is, within a specific organizational, cultural, and political context. In this way, the policy informs and frames how evaluation is carried out within a specific organization; that is, the evaluation policy serves as a *guiding mechanism* for evaluation practice. Indeed, if the policy does not serve this function, the other policy functions proposed by Trochim, communication, democratization, and accumulated learning, all of which are worth pursuing, cannot be realized in any meaningful manner.

We would also propose that an evaluation policy has to be written or made explicit in some other way. While recognizing that tacit, unstated rules and principles may influence evaluation practice (and that these unstated rules and their influence on practice should be understood), for the present purposes these are not considered formal evaluation policy. The purpose of evaluation policy is to inform how, when, in what way, and with what purpose to conduct evaluation within an organization. Accordingly, and to serve this purpose effectively, the evaluation policy has to be explicit.

Given this central role of policies for evaluation practice, good policy should be driven by a vision of how and in what way evaluation supports the organization (the aspired achievement) and identify the rules or standards for how evaluation may contribute to this vision (how to get there). The policy should consider why and toward what purpose the organization will need evaluative information and how and what way the organization will likely be obtaining evaluative information.

More specifically, and as formulated by the AEA Evaluation Policy Task Force, evaluation policies may guide an organization in terms of:

- *Evaluation definition.* How, if at all, is evaluation defined in an agency or in legislation? In such contexts, how is evaluation formally distinguished from or related to other functions such as program planning, monitoring, performance measurement, or implementation?
- *Requirements of evaluation.* When are evaluations required? What programs or entities are required to have evaluations? How often are evaluations scheduled? What procedures are used to determine when or whether evaluation takes place?
- *Evaluation methods.* What approaches or methods of evaluation are recommended or required by legislation or regulation, for what types of programs or initiatives?
- *Human resources regarding evaluation.* What requirements exist for people who conduct evaluations? What types of training, experience, or background are required?
- *Evaluation budgets.* What are the standards for budgeting for evaluation work?
- *Evaluation implementation.* What types of evaluation implementation issues are guided by policies? For instance, when are internal versus external evaluations required and how are these defined?
- *Evaluation ethics.* What are the policies for addressing ethical issues in evaluation? (AEA Evaluation Policy Task Force, 2009).

These dimensions reflect significant aspects of evaluation practice. As Trochim (2009) observes, ...developing well-informed evaluation policies that can guide evaluation practice may be the most important issue currently facing our field. [. . .] It touches virtually everything we think about or do in evaluation. (p. 14)

Evaluation policies, among other mechanisms, serve as an important connector between the collective body of knowledge in our field and the everevolving challenges that we face in our day-to-day evaluation practice. In particular, we would expect this to be true in the case of formal written evaluation policies. From this perspective alone, and as evaluation policies have become increasingly commonplace, there is a need for evaluation scholars to pay attention to evaluation

policies and the role they play in shaping evaluation practice—their purpose and aim and implications for evaluation practice deserve our attention. However, equally important to an understanding of the *policy–practice* relationship—the aim and scope of Trochim's inquiry—is a better understanding of the *theory–policy–practice* relationship, that is, the ways in which evaluation theory connects to policy, which then in turn translates into practice. This brings us to the theory–policy–practice nexus.

The Theory–Policy–Practice Nexus

The ways in which evaluation policies serve as a mechanism between evaluation theory and evaluation practice is little understood. We know a defining property of a policy in general, and in evaluation in particular, is that greater and lesser emphasis is placed on specific aspects of practice, so to result in a desired course of action (Trochim, 2009). This is also the case for evaluation theories—as theories of evaluation focus on describing what evaluation practice *should* be. The extent to which policies are informed by theory is important to understand, given the different but related purposes of evaluation theories and evaluation policies. Yet the conceptual and empirical work on the *theory–policy–practice* connection is limited.

While the set of papers comprising the *New Directions for Evaluation* volume centers on the *policy–practice* connection, not addressed substantively is the equally important connection between evaluation theory and evaluation policy—the *theory–policy* connection. One noteworthy exception is the following statement by Trochim (2009) on the relationship between evaluation theory and policy,

> The argument to be made here is that standards, guidelines, and theories become policies only if and when they are consciously adopted to guide decisions or actions about evaluation and when the organization institutes consequences for encouraging or enforcing them. (p. 17)

As Trochim (2009) goes on to illustrate, an evaluation theory, such as utilization-focused evaluation, becomes an evaluation policy "when an organization decides it is going to use this approach or adopt this theory in doing evaluations; failing to do so will have consequences" (p. 17). Underlying this argument is the framing of evaluation policies as "sticks" or "carrots," that is, as regulatory mechanisms defined by enforcement or encouragement of compliance (Vedung, 1999). This is a strong position.

While we agree with Trochim's emphasis on the importance of evaluation policies, we part ways with Trochim in a number of important ways. First, we hold that evaluation policies not only have to be "consciously adopted" but perhaps more importantly consciously *adapted* to the local political, cultural, and structural context of an organization. This is a seemingly trivial point, but, as we will go on to argue, a central purpose of evaluation policies is to contextualize evaluation theory. Whereas evaluation theory by nature, and as described above, is removed from the practical contingencies of local organizational contexts, evaluation policy serves—at least in part—to situate evaluation theory within the specific context of an organization. In this way, evaluation policy serves as a transformative mechanism between evaluation theory and practice.

Second, we take distance from the idea that evaluation theory becomes evaluation policy by way of enforcement or encouragement of compliance, that is, if and only if "the organization institutes consequences for encouraging or enforcing them" (Trochim, 2009, p. 17). The position we hold is that an evaluation policy becomes an evaluation policy when it is formulated and made explicit as such—consequences or no consequences. Moreover, and this is where we can advance our understanding of the theory–policy nexus, how and with what purpose evaluation theory is translated into evaluation policy may be broader and involve other mechanisms as well. That is, evaluation theory may not solely translate into regulatory stick-and-carrot

policies. For example, we might imagine, especially given the central role of policies for guiding practice, that evaluation theory becomes policy for the purpose of *enlightenment*— what Vedung (1999) would refer to as "sermons." Framed in this way, evaluation policies are perhaps much better understood as statements of intent or aids in thinking—a way to frame "a conversation" around evaluation. This is how policies may compel certain activities—by promoting certain principles, working logics, and so forth. The purpose of integrating evaluation theory into policy, then, is much more about clarifying key concepts and relatedly promoting a shared language and mind-set around evaluation within the particular political, cultural, and organizational context of an organization.

These observations are not raised as critiques—the purpose and role of evaluation policies, as well as the translation of evaluation theory into these policies, may in some cases work how Trochim describes them. The point to be made here is that the way in which evaluation theory translates into policy may also work in other ways. Accordingly, how we frame the evaluation theory–policy relationship holds import and in effect demands to be further unfolded and conceptualized. Advancing in this direction, toward sharpening our understanding of the ways in which evaluation theory is used in evaluation policies, examining illustrative examples of evaluation policies is a useful place to start.

Examining the Theory–Policy Nexus in Practice: Three Illustrative Examples

With the above observations as our backdrop, we now examine three evaluation policies as illustrative examples of the salience or influence of evaluation theory in evaluation policies. The evaluation policies we consider were selected to advance a discussion. We purposefully selected policies developed by major evaluation commissioners, representing different political, organizational, and cultural contexts. Accordingly, the observations made here are not to be viewed as grounding for empirical analysis or generalizations, but rather as a means to illustrate the conceptual points raised in the article.

Specifically, we examine the evaluation policies of the U.S. Administration for Children and Families (ACF), the Bill and Melinda Gates Foundation, and the U.S. Agency for International Development (USAID). These three commissioners are characterized by a significant and sustained involvement in evaluation, a long-term commitment to evaluation, and for significant investments in evaluation. The U.S. ACF is part of the Department of Health and Human Services—the largest federal commissioner of evaluation services (Lemire, Fierro, Kinarsky, Fujita-Conrads, & Christie, 2018), the USAID is the largest commissioner of evaluation in international development (Lemire et al., 2018), and the Bill and Melinda Gates Foundation is the largest commissioner of evaluation services among U.S. foundations (Kinarsky, 2018). Considered collectively, the three commissioners reflect markedly different yet equally significant involvements in evaluation. Each of the policies is available to the public and identified by the commissioners as their current and official "evaluation policy."

In what follows, we first briefly outline the content, purpose, and structure of each of the three evaluation policies, awarding particular attention to the ways in which each policy situates evaluation within the specific context of its commissioner. We also examine the consideration awarded methods, use, and valuing—three core dimensions of evaluation theory. In doing so, it is evident which dimensions of theory are most salient in the policies. Second, we reflect across the three evaluation policies to highlight salient similarities and distinctions. Table 1 offers a summary of what is presented in the text that follows.

Table 1. Main Features of Evaluation Policies and Evaluation Theory.

Feature	ACF (2012)	Bill and Melinda Gates Foundation (2013)	USAID (2011)
Structure	Structured around the five key principles that govern planning, conduct, and use of evaluation: rigor, relevance, transparency, independence, and ethics	Structured around five sections. The main section, Evaluation Design and Methods, is structured according to three types of evaluation. In addition, there are sections on purpose, context, and roles and responsibilities	Structured according to context, purpose of evaluation, roles and responsibilities, evaluation practices, evaluation requirements, and conclusion
Goal	To promote "rigor, relevance, transparency, independence, and ethics in evaluation"	To "help foundation staff and partners align their expectations in determining why, when, and how to use evaluation"	To provide, "clarity about the purposes of evaluation, the types of evaluations that are required and recommended, and the approach for conducting, disseminating, and using evaluations"
Intended audience	ACF staff and grantees	Gates foundation staff and partner organizations	USAID staff, partners, and stakeholders
Definition of evaluation	No specific definition provided	"Evaluation is the systematic, objective assessment of an ongoing, or completed intervention, project, policy, program, or partnership"	"Evaluation is the systematic collection and analysis of information about the characteristics and outcomes of strategies, projects, and activities as a basis for judgments to improve effectiveness, and timed to inform decisions about current and future programming"
Accountability and learning emphasis	Accountability over learning	Learning and accountability	Accountability for learning
Operational specificity	Low	Low–medium	High
Dimensions of theory			
Methods	– Traditional notions of rigor through research design (Campbellian) – Awards supremacy to experimental designs – Reflections on internal, external, and statistical conclusion validity	– Methodological pluralism/neutrality – Purpose-driven design (as opposed to one size fits all) – Rigor is defined as being "rigorous about the inferences [evaluators and partners] make and explicit about the assumptions they use to draw conclusions" (p. 4)	– Methodological pluralism/neutrality- "Evaluations to be undertaken in a manner that ensures credibility, unbiasedness, transparency, and the generation of high-quality knowledge" (p. 9) – Reflections on internal and external validity

Table 1. (continued)

Feature	ACF (2012)	Bill and Melinda Gates Foundation (2013)	USAID (2011)
Use	– Emphasis on stakeholder involvement and dissemination of findings, so that evaluation information is useful and relevant. However, to promote objectivity, ACF "protects independence in the design, conduct and analysis of evaluations" – No specification of different types of use or strategies for promoting use	– Emphasis on collaboration – Formative learning-oriented evaluations – Actionable evidence – Emphasis on roles and responsibilities for use	– Reflections on culture of accountability and leaning, leadership, and incentives (p. 2) – Integration of evaluation in program design and consultation with partners and beneficiaries – Use and dissemination internally and externally (wider community)
Valuing	"Evaluations should be designed to represent the diverse populations that ACF programs serve, and ACF should encourage diversity among those carrying out the work, through building awareness of opportunities and building evaluation capacity among underrepresented groups"	No mention	Local ownership of the evaluation. "The conduct of evaluations will be consistent with institutional aims of local ownership through respectful engagement with all partners, including local beneficiaries and stakeholders, while leveraging and building local evaluation capacity"

Note. ACF = Administration for Children and Families; USAID = U.S. Agency for International Development.

The Administration for Children and Families Evaluation Policy

The stated mission of the Administration of Children and Families is to "foster health and well-being by providing federal leadership, partnership, and resources for the compassionate and effective delivery of human services" (79 Fed. Reg. 51574–51575, 2014). In line with and support of this mission, the evaluation policy serves to reinforce the administration's commitment "to conducting rigorous, relevant evaluations and to using evidence from evaluations to inform policy and practice" (79 Fed. Reg. 51574–51575, 2014, p. 51574). Toward this end, the evaluation policy is structured around five guiding principles for ACF evaluations: rigor, relevance, transparency, independence, and ethics.

As stated in the opening of the policy, ACF views evaluation as but one of several types of evidence relevant for decision-making purposes within a "learning organization with a culture of continual improvement" (79 Fed. Reg. 51574–51575, 2014, p. 51574). Other types of evidence include descriptive research studies, performance measures, and financial data, among others. The five guiding principles are considered applicable across these different types of evidence. The evaluation policy primarily places emphasis on *methods* and to some extent *use* in its presentation of the guiding principles. The presence of *valuing* is limited.

Grounded in the Campbellian (a methods theorist) model of validity, the first principle of rigor is defined in terms of internal and external validity (see Shadish, Cook, & Campbell, 2002). Aligned with this emphasis, the primary focus of the ACF principle of rigor concerns "assessing" and "isolating" the "effects" of programs, whereby determining the least biased net-effect estimate is central (79 Fed. Reg. 51574–51575, 2014, p. 51574). For "causal questions," the policy states, "experimental approaches are preferred. When experimental approaches are not feasible, high-quality quasi-experiments offer an alternative" (p. 2). While the policy states that "rigor is not restricted to impact evaluations" (79 Fed. Reg. 51574–51575, 2014, p. 51574), the emphasis on the latter illustrates well the overall emphasis of the policy on traditional notions of rigor and social accountability, and in its stated preference for experimental approaches, on traditional design-oriented approaches for determining "what works."

The emphasis on methods is also evident in the principles of *transparency* and *independence*. In relation to transparency, the importance of evaluation reports providing descriptions of strengths and weaknesses of the methods used (i.e., internal validity) as well as the generalizability (external validity) of findings produced by the evaluation is highlighted. In tandem with this emphasis on methods, the principle of independence affirms that "independence and objectivity are core principles of evaluation," why "it is important to insulate evaluation functions from undue influence and from both the appearance and the reality of *bias*" (p. 3, our emphasis). Taken collectively, the principles of rigor, transparency, and independence exemplify the overall orientation toward methods in the ACF policy.

Reaching beyond methods, the use dimension of evaluation is primarily reflected in the guiding principle on *relevance*. According to the policy, relevance is achieved by aligning evaluation priorities with "Congressional interests" as well as "the interests and needs of ACF, Health and Human Services (the agency in which ACF is under), and Administration leadership; program office staff and leadership; ACF partners such as states, territories, tribes, and local grantees; the populations served; researchers; and other stakeholders" and by disseminating findings in ways that are "relevant and useful" for these stakeholders (79 Fed. Reg. 51574–51575, 2014, p. 5157479 Fed. Reg. 51574–51575, 2014, p. 51574). In addition, the policy states that evaluations will be more "feasible and useful" when designed in tandem with early program planning and development. The operational specificity on how one might achieve these outcomes is not stated explicitly in the policy. No distinction is made between different types of use or of the specific strategies one might use to achieve these outcomes. The emphasis on use is pragmatic in purpose, that is, it is focused on

enhancing the relevance and usefulness of the evaluations for policy decision makers and their decision-making needs.

Another dimension of use, more specifically evaluation capacity building, is also mentioned in the policy: "evaluations should be designed to represent the diverse populations that ACF programs serve, and ACF should encourage diversity among those carrying out the work, through building awareness of opportunities and building evaluation capacity among under-represented groups" (79 Fed. Reg. 51574–51575, 2014, p. 51574). Here, we see evidence of a concern for the importance of conducting evaluations that are culturally relevant and responsive, perhaps even transformative in purpose. This indicates a concern for ensuring that underrepresented groups have a role in the evaluation process—not only as stakeholders but also as those who are commissioned to conduct evaluation studies. These practices are informed by some of the theoretical ideas that are salient on the valuing branch of evaluation theory.

In sum, what we see is an evaluation policy firmly grounded in postpositivism, with a secondary pragmatic concern for providing useful information for policy makers. There is also an aim to be relevant to traditionally underrepresented groups and to accomplish this goal, there is a recognition of the need to build great evaluation capacity in communities that are often studied in evaluation studies. This is a policy is best characterized as having a strong methodological focus, grounded in more traditional notions of impact evaluation, validity, and research designs. The policy weighs accountability over learning, impact over process. The principle of rigor stands strongest both in terms of page space and placement (first principle). In line with this emphasis, the operational specificity provided across the five principles is generally limited, but more prescriptive procedural detail is provided for the methods aspect of the policy, awarding methodological supremacy to experimental design over other evaluation designs.

The Bill and Melinda Gates Foundation Evaluation Policy

The Gates Foundation policy is comprised of five main sections including general Introduction and Conclusion sections and three main sections (Our Strategies and Evaluation, Evaluation Design and Methods, and Evaluation Roles and Responsibilities, respectively). These latter sections elaborate more directly on evaluation issues that are of central concern to the foundation. The stated purpose of the policy reads:

> Our evaluation policy is intended to help foundation staff and partners align their expectations in determining why, when, and how to use evaluation. More specifically, the policy encourages foundation teams to be more transparent, strategic, and systematic in deciding what and how to evaluate. (p. 1)

This establishes the policy as one that describes both the collaborative management and practice of evaluation, with an eye toward principles associated with use theories that prioritize the decision-making needs of multiple organizational decision makers. By extension, the policy also aims "to integrate evaluation into the fabric of our work, achieve early alignment with our partners about what we are evaluating and why and generate evidence that is useful to us and our partners as we move forward" (p. 1). The general emphasis is on use and learning—which echoes the overall mission of the Gates foundation.

In the most substantive section of the policy, *Evaluation Design and Method*, use principles are also evident, describing a goal for evaluation as wanting "our evaluation efforts to be designed for a specific purpose and for specific intended users" (p. 4). This is very reminiscent of Patton's (1978; a use theorist) intended *use by intended users* principle. Toward achieving this goal, the policy advocates for a "fit-for-purpose" approach to evaluation, which counters the "one-size-fits-all" approach to evaluation and serves to broaden the range of approaches and methods to be used in

evaluations (p. 4). The emphasis on methodological pluralism echoes Rossi's (a methods theorist) writings on conducting tailored evaluations (Rossi, Lipsey, & Freeman, 2004).

By extension of the fit-for-purpose approach, the policy also identifies three different types of evaluation typically conducted in the foundation: (1) Evaluations to understand and strengthen program effectiveness (program improvement); (2) Evaluations to test the causal effects of pilot projects, innovations, or delivery models (program effectiveness); and (3) Evaluative activities to improve the performance of institutions or operating models (organizational improvement; pp. 4–6). The different types of evaluation are defined in terms of purpose and are accompanied by descriptions of scenarios where they might be used, with a set of salient design considerations. This latter guidance provides operational specificity for the methods and approaches relevant for each type of evaluation. Yet, in advancing the different types of evaluation, the policy remains methodologically neutral in recognition that the "diversity of partners and areas of focus precludes [the policy] from promoting only certain types of evaluation evidence as acceptable for decision making," suggesting that credible evidence is determined—at least in part—by stakeholders (p. 4). In line with the emphasis on collaboration and use of evaluation, the notion of rigor invoked in the policy is applied to "decision making on when and how to use evaluation and the types of evaluation to implement," which represents a stark departure from traditional notions of (methods oriented) rigor and aligns more firmly with Wholey's sequential purchase of information for decision-making (Wholey, Hatry, & Newcomer, 2010, p. 2). This positioning of rigor reflects the overarching aim of the policy to promote more systematic and transparent decision-making about when and how to evaluate. Moreover, the policy's emphasis on development of "actionable evidence" is in line with the commitment to provide credible and useful knowledge for stakeholder needs.

Reasserting this collaborative orientation, evaluation is viewed as a collaborative "learning tool" that can be used for "testing innovation, making improvements, and understanding what works and why to learn quickly from failure and replicate success" (p. 2) for both foundation staff and partners. The policy also mentions the importance of evaluation capacity building as part of supporting global programs (p. 9), though the emphasis on the latter is limited.

In sum, the Gates Foundation policy is best characterized as having a strong emphasis on methodological pluralism and how to decide between different types of evaluation (purpose-driven evaluation). In this way, the presentation of the different types of evaluation and related methods combines high specificity with a descriptive orientation—no strong prescriptions for practice are provided. Also notable is the emphasis on use of evaluation as part of collaborative, organizational learning (as opposed to accountability). The emphasis on valuing is limited.

The USAID

The USAID policy is the lengthiest policy (12 pages) of the three examined. The stated purpose of the policy is to provide "clarity about the purposes of evaluation, the types of evaluations that are required and recommended, and the approach for conducting, disseminating, and using evaluations" (p. 2). The general emphasis is on accountability and learning. The policy is structured according to context, purpose of evaluation, roles and responsibilities, evaluation practices, evaluation requirements, and conclusion. Thus, the policy provides both management and procedural guidance.

Postpositivist thinking is salient in this policy. It is stated, "Evaluations will be undertaken in a manner that ensures credibility, unbiasedness, transparency, and the generation of high-quality knowledge" (p. 9). While there is a general emphasis on methodological pluralism/neutrality, whereby the best method/approach is the determined by the purpose and context of the evaluation, experimental designs are highlighted as the best design for impact evaluation.

There is also an emphasis on use throughout the policy. Reflections are offered on the importance of a culture of accountability and learning, leadership, and incentives (p. 2), as well as on integration

of evaluation in program design and consultation with partners and beneficiaries, and a long section on the roles and responsibilities of those involved in the evaluation. The reach of the intended use is both internal and external (wider community). The latter is reflected in a requirement that final evaluation reports and data sets be submitted and made available in the Development Experience Clearinghouse.

Valuing is reflected in the policy in limited ways. Most significant, under the evaluation practices section of the policy, there is subheading that addresses the local ownership of the evaluation. Toward this end, it is stated that:

> The conduct of evaluations will be consistent with institutional aims of local ownership through respectful engagement with all partners, including local beneficiaries and stakeholders, while leveraging and building local evaluation capacity . . . . USAID will place priority within its sectoral programming on supporting partner government and civil society capacity to undertake evaluations and use the results generated. (p. 8)

In line with this thinking and in the same section of the policy that prescribes the programming priority to be aligned with local norms, cultures, and customs, the salience of objectivity is also emphasized, stating: "To the extent possible, evaluation specialists with appropriate expertise from partner countries, but not involved in project implementation, will lead and/or be included in evaluation teams" (p. 8). In this way, the methods and valuing dimensions of the policy overlap, advocating for objectivity in a culturally sensitive way.

In sum, this policy is concerned mostly with methods that are akin with those philosophies and principles associated with traditional notions of rigor and accountability, which are represented on the methods branch of the evaluation theory tree. There is certainly also an emphasis on use, and evaluation as a tool not only for accountability but also for decision-making. While limited concerns about local ownership of the evaluation do reflect some theoretical ideas associated with the valuing dimension of evaluation, the policy also outlines the expectation, and provides specificity on implementation, for managing and carrying out the evaluation policy guidance.

Crosscutting Insights

As expected, the above discussion reveals both similarities and differences across the three policies. First, we acknowledge that the policy writers observably drew upon theory in the development of these policies. We know this not only through the analysis done here but also through discussions and presentations at conferences and other meetings. What is interesting to us is what theoretical ideas are more and less salient in the read of these policies.

The policies share an overarching commitment to evaluation-informed decision-making, citing the important role of evaluation in terms of "informing policies and practice" (79 Fed. Reg. 51574–51575, 2014, p. 51574), "decision making about how to optimize scarce resources for maximum impact" (Gates, p. 1), or as a "means through which [USAID] can obtain systematic, meaningful feedback about the successes and shortcomings of its endeavors" (USAID, p. 1). Evaluation is even viewed as "fundamental to the Agency's future strength" (USAID, p. 1). Positioning evaluation within organizational decision-making is a central aim in all three policies.

The policies, however, differ markedly in their specific purposes and scope. Whereas one emphasizes accountability over learning (ACF), another emphasizes learning over accountability (Gates), while a third places emphasis on learning for accountability (USAID). Whereas one aims to promote rigorous evaluation (ACF); another aims to align expectations in relation to why, when, and how to use evaluation (Gates); while a third aims to promote clarity about the purposes and different uses of different types of evaluation (USAID). These fundamental differences in many ways reflect the nature of the organizations and the overall purpose and orientation of the policies within these.
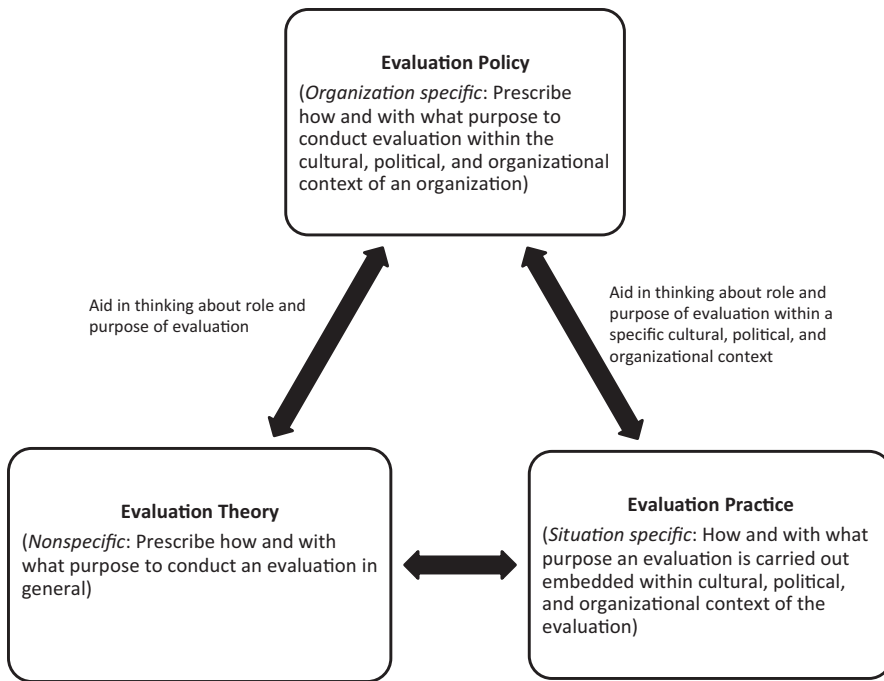
Figure 1. Theory-policy-practice relationship in evaluation.

In terms of evaluation theory, we found distinct traces of evaluation theories in each of the three policies. The methods branch was awarded much more attention across all three policies, as compared with use, and the valuing dimension only received scant attention. However, and of particular salience in the present context, the policies also displayed noticeable variations in their emphasis on and interpretation of methods, use, and valuing—again, corresponding to the overall aim and purpose of the evaluation policy within their respective organizational context. To illustrate, the three policies all share a commitment to rigorous evaluation, although their definitions of rigor differ. Whereas the ACF policy defines rigor in Campbellian terms, referencing traditional notions of internal, external, and measurement validity; the Gates Foundation policy places emphasis on rigorous decision-making around when, why, and how to evaluate, which is more aligned with Wholey's sequential purchase of information for decision-making (Wholey et al., 2010). Situating evaluation as a way of learning about effectiveness, the USAID policy places emphasis on "credibility, transparency, and the generation of high-quality information and knowledge" (p. 9). This reminds us of Weiss's (1998) writings on the importance of generating rigorous credible evidence for decision-making.

In a similar vein, the important role of the use of evaluation is also evident across the policies, although, in different ways. In the ACF policy, where the focus on use is secondary to that of methods, attention is awarded to stakeholder involvement and dissemination of findings that is useful and relevant. There is no specification of different types of use or specific strategies for promoting use, however. In the Gates Foundation policy use is quite salient, with an emphasis on collaboration through primarily formative, learning-oriented evaluations and for the purpose of generating actionable evidence. Finally, the USAID policy can be situated somewhere between the two other policies with a balance of learning and accountability, while it also offers reflections on the importance of establishing a culture of accountability and learning, leadership, and incentives (p. 2). Moreover, the policy seeks to promote integration of evaluation in program design and

consultation with partners and beneficiaries and places a heavy emphasis on the different roles and responsibilities toward these aims. The reach of evaluation use is both considered internally and externally (wider community).

These differences in emphasis and interpretation of the use and methods dimensions are particularly interesting in the present context because they reflect the role and purpose of evaluation within each of the three organizations—that is, the organizational context. To illustrate, the role of evaluation in ACF, at least as described in the policy, is to assess the effectiveness and performance of programs. As such, the emphasis of the policy is consistently on the importance of enhancing these principles in ACF evaluations and in effect on methods and accountability. In marked contrast, and again on the basis of the description provided in the policy, evaluation in the Gates Foundation appears to serve a much broader role and purpose, including promoting program improvement, documenting program effectiveness, and improving the performance of partner institutions. Accordingly, the emphasis of the evaluation policy—in both structure and content—is on how to align these different purposes with evaluation designs and approaches. The subject matter is not how to improve evaluation designs in a direct sense but rather how to approach the design of evaluations. In addition, the collaborative nature of the Gates foundation results in a policy that is intended for Gates Foundation staff and for partnering agencies.

Finally, the broader context of the organizations also appears to influence their evaluation policies. For instance, the emphasis of the ACF policy, which is under the federal government, is oriented much more firmly toward accountability. Similarly, the USAID policy, emphasizing accountability and to some extent learning, is also a public-funded entity. In contradistinction, the Gates Foundation policy, while mentioning both learning and accountability, places a heavy emphasis on organizational learning. These differences may also reflect the core principles and philosophy of the organizations. To illustrate, the Gates Foundation policy clearly states its connection with "the foundation's core values: collaboration, rigor, innovation, and optimism" (p. 2).

It also deserves mention that the evaluation policies often state the importance of different approaches and methods, yet generally stop short of mentioning specific approaches (e.g., utilization-focused evaluation, empowerment evaluation). The policies attend and commit to use and decision-making, yet they offer no attention to the different kinds of use and award no attention to the conditions and strategies that might promote use (simply consider the salience of Weiss's work on use and conditions for use). While the policies give some mention to valuing, they stop short of engaging with the rich body of theory on this topic, including the theoretical work on value-engaged evaluation (Greene, DeStefano, Burgon, & Hall, 2006) and democratic evaluation (House & Howe, 1999), among other salient contributions in this area. Accordingly, there is an unrealized potential for evaluation theory in the examined evaluation policies.

To conclude this section, we note that to some readers, our observations about the presence of evaluation theory in these policies are perhaps apparent. However, their implications for evaluation practice are potentially very real and for that reason worth our attention in an effort to better understand how evaluation policy serves as a mechanism between evaluation theory and practice. We turn to this topic in what follows.

Strengthening the Theory–Policy Relationship in Evaluation

In this article, we argue three points. First, that evaluation theory is removed from the practical contingencies of local organizational contexts. Second, that evaluation policy helps to operationalize evaluation within the context of an organization. And third, that evaluation policy may serve as an important translational mechanism between evaluation theory and evaluation practice. We depict this relationship in Figure 1.

As seen in Figure 1, we maintain that evaluation policies help to ensure that evaluation practice is aligned with an organization's specific political, organizational, and cultural values, norms, and beliefs about evaluation. In this way, evaluation policy serves as a transformative mechanism for translating and perhaps more importantly situating evaluation theory within the organizational, political, and cultural context of the organization. This is important as evaluation theory is intended to provide more general principles related to the role and purpose of evaluation; and more specifically, how theoretical principles related to methods, use, and valuing (Christie & Alkin, 2012) should influence the conduct of the evaluation so to articulate the theory's primary purpose (as held central to the theory). Evaluation practice is the articulation of an evaluation policy in the context of a specific evaluation study. Of course, evaluation practice can also be an articulation of evaluation theory, as many have argued for decades is a primary purpose of evaluation theory. We reason, however, that theory may be better integrated, operationalized, and more applicable to practice when translated through policy. Therefore, the integration of theory into policy should be intentional and explicit and focus on adaption in service of *translating* theory into practice.

For the purposes of enlightenment, evaluation theory may strengthen evaluation policies in several distinct ways. First, and informed by Shadish (1998), evaluation theory may serve to clarify key concepts and, relatedly, support a shared language of evaluation across diverse organizations and policies. As reflected in the three policies above, central—yet often confused—evaluation terms were common across the policies, including terms such as *use, rigor, theory*, to name but a few. Anchoring these terms more firmly and explicitly in existing evaluation theory would serve to specify and align the terms across the policies as well as to elevate discussions and debates about key evaluation concepts and issues. To illustrate, the Gates Foundation policy awards sustained emphasis on the use of evaluation across and between multiple stakeholder organizations. In this way, evaluation theory may clarify key concepts (use, rigor, among others) and may also serve to describe different approaches to (models for) evaluation, including utilization-focused evaluation, empowerment evaluation, value-engaged and democratic evaluation, among others (even if the policy refrains from a prescriptive stance on these).

Another benefit of integrating evaluation theory more firmly into evaluation policies is that it promotes a better understanding of the contingency of evaluation theory. Being explicit about which theories have been infused into policies will provide the evaluation community with a new and different opportunity to advance our thinking on which theories, elements of theories, or theoretical principles are best used under particular conditions. While these discussions have ensued for many years, having these discussions in the context of evaluation policy—a key driver of practice—will provide the field with an opportunity to think about the role of theory in a new way. For example, we might explore discussions on the extent to which evaluation theory should be infused into policy, which aspects of evaluation theory, if any, should be infused, and so forth. In extension, this theory integration should advance our ability to examine how evaluation theory, by way of evaluation policy, may influence practice. And while research on evaluation tells us that it is highly unlikely that any one single theory is used to guide practice, if a single theory is explicitly used by an organization to guide the formulation of an evaluation policy (e.g., utilization focused evaluation), then, as Trochim (2009) explains, the evaluation policy can be examined as a direct conduit for understanding that particular theory in practice.

Toward advancing our understanding of the theory–policy–practice nexus, the present article motivates more questions than it answers, hopefully stimulating and paving the way for further avenues of inquiry. One question that emerges revolves around the nature and dynamics of the theory–policy–practice nexus. On this topic, it would be insightful to know whether theory was purposefully infused into evaluation policies. As Weiss noted, any type of policy development is a process (Weiss, 1998). To establish a good policy, one that will ensure (as far as possible) that good outcomes emerge from it, much effort is required. The policy development process is dynamic and

policy environments, both at the macro (i.e., federal) and micro (i.e., organizational) levels, are full of complexities, usually involving a diverse range of actors with different perspectives. It is, there-fore, very unlikely that circumstances would permit a linear, rational decision-making process. Strategies for infusing evaluation theoretical ideas into evaluation policy should be no different from infusing ideas into any other policy-making process. There are known techniques for doing this well that include educating policy makers and organizational leaders through publications and seminars; establishing task force groups where those who are expert in evaluation theory are engaged in ongoing interactions with policy makers and advisors; a good network of relationships between those who are expert in evaluation theory both inside and outside of government (or in the case of organizational policies, the organization). The AEA is working to accomplish this through its evaluation policy task force. While the policy literature suggests that high levels of interaction between the evaluation experts and policy makers should lead to good evaluation policy outcomes, we assume that this will not necessarily be the case in all circumstances.

Trochim (2009) suggests that many of the debates that have been thought to be theoretical are rather debates about evaluation policy. Lets take for example the controversy around U.S. govern-ment policies that give preference to proposals for evaluation studies that utilize randomized con-trolled trial designs. Central to the thesis presented in this article, we argue that these are indeed theoretical debates—debates that center on what constitutes credible evidence, that are grounded in the philosophies of science—that have influenced policy and, in turn, the type of evaluation practice that is seen as rigorous and worthy of federal support. Thus, it is evident that we need to better understand how and under what conditions evaluation theory becomes influential in the policy development process. By extension, a better understanding of how evaluation theory is used to inform policies (if at all) might also advance more reflective practice around how and for what purpose evaluation theory is applied in policies.

While we argue that for an evaluation policy to be a policy it needs to be formal, explicit, and written, we acknowledge that organizations operate using informal evaluation policies that influence practice. The consequential role of informal or implicit evaluation policies on evaluation practice has been identified in a study of evaluation policy at the Robert Wood Johnson Foundation (Dillman & Christie, 2017). The illustrative analysis of three policies presented in this article obviously focus on explicit, formal published policies. Yet, we know that to fully understand the connections between evaluation theory and policy the more implicit policies and practices that an organization operates under when conducting evaluation must also be understood. We must also investigate the extent to which theories outside of evaluation shape the positions presented in evaluation policies. It is quite plausible, for example, that organizational learning theories or theories of social justice may inform an organization's evaluation policy.

Returning to the main topic of the present article, we begin to see how the theory–policy relationship may be central to evaluation. To understand why evaluation is being practiced as we observe practice to be and the extent to which theory influences or guides this practice is not simply a question of which theories practitioners prefer, understanding the contexts in which practitioners' work, or the traditions in which practitioners were trained (Christie & Masyn, 2008). Rather, it is a confluence of factors, including the very important determinant of evaluation policy. As such, and if we are to more fully understand the theory–practice connection, the role of evaluation policies will have to be considered. This mediated relationship may be more explanatory than the direct theory–practice relationship we have been concerned with for decades. Ultimately, we are still concerned with the nexus of theory and practice but are arguing here that there are two paths by which this connection might occur—directly and one that is mediated by evaluation policy. Our hope is that the present contribution serves as a motivating stepping stone toward more awareness, discussion, and analysis of the theory–policy–practice nexus.

Declaration of Conflicting Interests

Funding

References

79 Fed. Reg. 51574–51575 (August 29, 2014)

Administration for Children and Families. (2012). *Evaluation policy*. Retrieved September 18, 2017, from https://www.acf.hhs.gov/evaluation-policy

Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots*. Thousand Oaks, CA: Sage.

Bill and Melinda Gates Foundation. (2013). *How we work—Evaluation policy*. Retrieved September 18, 2017, from https://www.gatesfoundation.org/How-We-Work/General-Information/Evaluation-Policy

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, *97*, 7–36. doi:10.1002/ev.72

Christie, C. A., & Alkin, M. C. (2012). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots* (2nd ed.). Thousand Oaks, CA: Sage.

Christie, C. A., & Masyn, K. E. (2008). Profiles of evaluators reported practice: A latent profile analysis. *Canadian Journal of Program Evaluation*, *23*, 225–254.

Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, *34*, 7–22.

Diesing, P. (1966). Objectivism vs. subjectivism in the social sciences. *The Philosophy of Science*, *33*, 124–133.

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*, 60–79. doi:10.1177/1098214016642864

Fetterman, D. M. (1994). Steps of empowerment evaluation: From California to cape town. *Evaluation and Program Planning*, *17*, 305–313.

Greene, J. C., DeStefano, L., Burgon, H., & Hall, J. (2006). An educative, values-engaged approach to evaluating STEM educational programs. *New Directions for Evaluation*, *109*, 53–71.

House, E., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.

Kinarsky, A. R. (2018). The evaluation landscape: U.S. foundation spending on evaluation. *New Directions for Evaluation*, *160*, 81–96. doi:10.1002/ev.20342

Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018). The U.S. federal evaluation market. *New Directions for Evaluation*, *160*, 63–80. doi:10.1002/ev.20343

Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverly Hills, CA: Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Thousand Oaks, CA: Sage.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation—A systematic approach*. Thousand Oaks, CA: Sage.

Scriven, M. (2004). *Practical program evaluation: A checklist approach*. Claremont Graduate University Annual Professional Development Workshop.

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

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.

Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. *New Directions for Evaluation*, *123*, 13–32.

Trochim, W. M. K., Mark, M. M., & Cooksy, L. J. (Eds.), (2009). Evaluation policy and evaluation practice [Special issue]. *New Directions for Evaluation*, 123.

U.S. Agency for International Development. (2011). The USAID evaluation policy—Evaluation learning from experience. Retrieved September 18, 2017, from https://www.usaid.gov/sites/default/files/documents/1870/USAIDEvaluationPolicy.pdf

Vedung, E. (1999). Public policy and program evaluation. *Administrative Science Quarterly*, *44*, 433–436. doi: 10.2307/2667008

Weiss, C. H. (1998). *Evaluation research: Methods of assessing program effectiveness*. Upper Saddle River, NJ: Prentice Hall.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass.