

Evaluation policy and organizational evaluation capacity building: A study of international aid agency evaluation policies

Hind Al Hudib  | J. Bradley Cousins 

Centre for Research on Educational and Community Services, University of Ottawa, Ottawa, Ontario, Canada

Correspondence

J. Bradley Cousins, Centre for Research on Educational and Community Services, 136 Jean-Jacques Lussier, VNR 5002, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada.
Email: bcousins@uottawa.ca

Abstract

Research and theory on evaluation capacity building (ECB) and organizational evaluation capacity have been developing at a good pace over the past decade. On the other hand, there is a paucity of research on the nature and consequences of organizational evaluation policy. Evaluation policies are developed and implemented ultimately to inform and shape evaluation practice and its consequences. It is therefore natural to consider the interface between evaluation policy and ECB. The present exploratory descriptive study examines 52 evaluation policies from bilateral and multilateral aid agencies to explore connections between evaluation policy content and ECB principles and considerations. The results shed light on some interesting relationships; they are discussed in terms of evaluation use, evaluation purposes, and organizational leadership. The study also resulted in a revision of Trochim's (2009) definition of evaluation policy and a refinement and expansion of his eight-category taxonomy. Implications for ongoing inquiry are considered and practical implications are offered to organization members and evaluation policy developers.

INTRODUCTION

In recent years, many governmental, para-governmental, and not-for-profit organizations and agencies have established or revised their evaluation policies to meet their accountability requirements, manage their resources, and guide their evaluation practices. However, searches of the evaluation literature for studies on the nature of evaluation policy, its consequences, and the antecedent conditions affecting it return very limited results.

The authors thank the Data Foundation for funding for open access.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *New Directions for Evaluation* published by American Evaluation Association and Wiley Periodicals LLC.

At the same time, research and scholarship in organizational capacity for evaluation have increased considerably over the past several years (Labin et al., 2012; Nelson et al., 2021). If evaluation policy is expected to influence evaluation practice at the organizational level, a close examination of the relationship between evaluation policy and organizational evaluation capacity building (ECB) makes good sense. Yet research with such a focus is virtually non-existent.

This chapter addresses this gap by presenting the findings from a comprehensive content analysis of evaluation policy documents from different international aid organizations. This exploratory descriptive analysis is intended to shed light on the extent to which contemporary evaluation policies are conducive to advancing organizational ECB interests.

EVALUATION POLICY

In 2009, Trochim and colleagues published an impressive collection of chapters that addressed a range of issues in *New Directions for Evaluation (NDE)*. Trochim (2009) included the following definition of evaluation policy: “An evaluation policy is any rule or principle that a group or organization uses to guide its decision and actions when doing evaluation” (p. 16). According to him, evaluation policy can specify a range of requirements related to evaluation and is critically important to organizational evaluation practice. In his words, evaluation policy

...is more important than getting our methods right, or concerns about validity. It is more important than the ethics of evaluation. It is more important than making evaluation participatory, or using it for empowerment. Why? Because evaluation policy encompasses all of those things and more. It touches virtually everything we think about or do in evaluation. (p. 14)

Typical questions that are addressed by evaluation policies include how evaluation should be done, what resources are needed, who is responsible for which tasks, what types of expertise are required, what are the expectations for evaluation, and how evaluations should be conducted (Mark et al., 2009).

Trochim (2009) laid out a theoretical framework for evaluation policy and several ways to further our knowledge about it. His emphasis reflects a shift in focus from customary evaluation routines to the effects that they have on evaluation practice. However, despite his efforts, a review of the relevant literature shows that evaluation policy is not integral to the discourse on how to enhance organizational capacity for evaluation.

Trochim’s (2009) eight-category “*draft* taxonomy of evaluation policies” (p. 20, emphasis added) highlighted several salient dimensions of evaluation policy and identified the inter-relationships between such policies and practices. These eight categories are as follows:

- evaluation goal (i.e., description of goals of evaluation in the organization or context),
- evaluation participation (i.e., who can be involved in evaluation, how and when),
- ECB (i.e., development of capacity to support evaluation activities),
- evaluation management (i.e., guidance for time resources, resource allocation, evaluation scheduling),
- evaluation role (i.e., responsibilities for evaluation),
- evaluation process and method (i.e., recommended or required evaluation approaches and methodologies),
- evaluation use (i.e., how evaluation results will be used), and

- meta-evaluation (i.e., periodic assessment of the implementation, quality, and utility of evaluation).

Leeuw (2009) used this framework to assess the state of evaluation policy in the Dutch government and identified some notable gaps. His study confirmed that Trochim's framework has great potential for describing and assessing evaluation policy, which would be of considerable usefulness in focusing efforts to develop organizational evaluation capacity.

Both Trochim (2009) and Cooksy et al. (2009) call attention to the need for empirical validation of salient dimensions of evaluation policy such as those in Trochim's taxonomy. Cooksy et al. suggest that such categories be checked against existing evaluation policies to determine the extent to which they are comprehensive and representative. Such verification would represent a first step toward refining "our understanding of the domain of evaluation policy" (p. 104). Given evaluation policy's overt intention of guiding evaluation practice, such analysis may also shed light on valued outcomes including organizational evaluation capacity. To what extent are evaluation policy elements aligned with or conducive to organizational intentions to build evaluation capacity? Conversely, are evaluation policy elements somehow at odds with an organizational ECB agenda?

ORGANIZATIONAL ECB

Considerable research on ECB has followed from the publication of an earlier volume of NDE featuring some ground-breaking work edited by Compton et al. (2002). In an early chapter, the authors define ECB as a

... context-dependent, intentional action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/programs/sites. (Stockdill, Baizerman, & Compton, 2002, p. 8)

Since 2002, a number of conceptual frameworks that emphasize different aspects of ECB have been published (e.g., Labin et al., 2012; Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008), which speaks to the relative maturity of the field. Labin et al. (2012) noted a high degree of consistency between empirical and theoretical scholarship on ECB at the individual (attitudes, knowledge, and behaviors) and organizational (practices, leadership, mainstreaming, and resources) levels. This consistency, they argued, implies that the field is ready for common measures, stronger designs, and more systematic reporting.

Yet others have argued that, while there has been considerable focus on the role of ECB, not much attention has been directed to the organizational evaluation capacity construct itself (Cousins et al., 2008; Nacarrella et al., 2007; Nielsen et al., 2011). Cheng and King (2017) observed that many studies have either blurred or failed to attend to the distinction between evaluation capacity and ECB, the process leading to it. Further, Cousins and Bourgeois (2014) noted, "*evaluation use* has been underexplored and underemphasized in theory, research, and practice concerning organizational evaluation capacity" (p. 10, emphasis added). Cousins and associates integrated organizational capacity to *do* evaluation and capacity to *use* it in a comprehensive framework (Cousins et al., 2008).

The Cousins and associates (2008) framework was used by Cousins and Bourgeois (2014) to guide a significant multiple-case study of eight organizations that yielded insights into the factors that influence ECB such as: administrative commitment and leadership; an organizational culture of learning; and the existence of resident evaluation-related

expertise. Organizational strategies to foster ECB included structural configurations and partnering arrangements. Finally, positive ECB was associated with a cycle of experiencing successful use and valuing of evaluation, direct ECB activities (e.g., workshops, resource augmentation), evaluation tools and mechanisms, and external stakeholder involvement and relationship building.

In tandem with this research, Bourgeois and Cousins (2008, 2013) developed and validated a profile-based instrument to assess the principal dimensions of organizational evaluation capacity. On the one hand, organizational capacity to implement evaluation is a function of human resources (e.g., staffing, technical skill), resources (e.g., infrastructure), and evaluation planning and activity (e.g., evaluation plan, information sharing, external support). On the other hand, organizational capacity to use evaluation arises from evaluation literacy (e.g., focus on results, involvement/participation), decision making (e.g., management processes, support), and learning benefits (use of findings, process use).¹ This evidence-based framework has been successfully validated for research and practice in governmental and other public sector organizations (Bourgeois & Cousins, 2008; Bourgeois et al., 2016; Bourgeois et al., 2013; Buetti et al., 2019).

The present study is an exploratory descriptive analysis of evaluation policies with an eye to understanding their implications for organizational ECB. While this inquiry is unable to generate firm cause-and-effect assertions, its speculative conclusions provide a good basis from which ongoing inquiry can proceed. The study is guided by a conceptual framework described in the next section. After describing the framework, we briefly share our methods and the findings from the descriptive analysis. The paper concludes with a discussion of issues for consideration and implications for research and practice.

CONCEPTUAL FRAMEWORK

What is the role of evaluation policy in enhancing organizational evaluation capacity? Integrating the aforementioned frameworks for organizational evaluation capacity (Bourgeois & Cousins, 2008, 2013) and evaluation policy (Trochim's, 2009) provides useful guidance to answer this question. Figure 1 represents our conceptual approach – through this Figure we see how aspects of organizational evaluation capacity are potentially shaped, reinforced, or impeded by salient evaluation policy dimensions. Ultimately, an organization's growth in capacity could eventually shape ongoing changes and evaluation policy revisions.

In the current qualitative study, our focus is on understanding the components of an evaluation policy, the ways in which evaluation policies are conceptualized within organizations, and how they are designed with a view to positively influence the evaluation function, in particular organizational capacity for evaluation. The elements of Trochim's taxonomy can be applied in different ways, ways that are more or less conducive to organizational capacity to do and use evaluation. Our descriptive content analysis helps shed light on this connection.

METHOD

Sample

Working from Trochim's definition of evaluation policy, we conducted manual searches of the websites of a wide range of (mostly Western) bilateral and multilateral aid agencies for evaluation policy documents. All websites had to include up-to-date English-language information and documents about evaluation policy and practices. Most of the documents

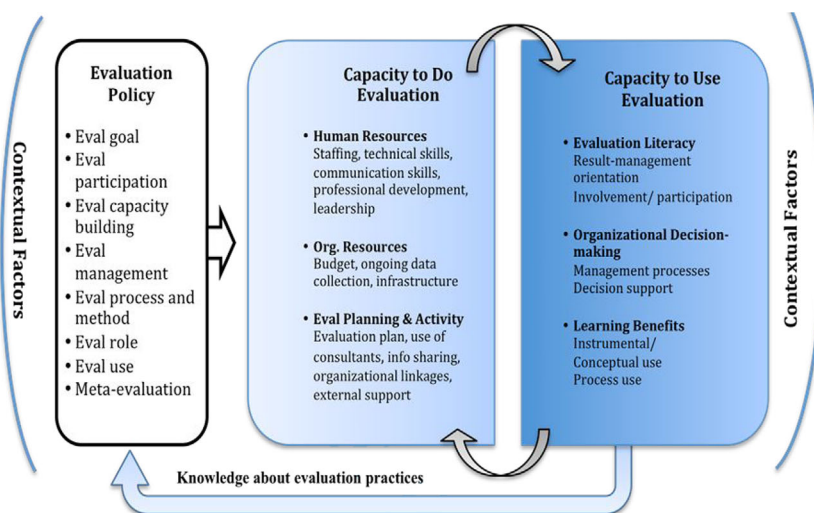


FIGURE 1 Conceptual framework describing the relationship between evaluation policy and organizational capacity to do and use evaluation

we found were labeled “evaluation policy” although a few that qualified carried various labels such as “guidelines,” “handbook,” “manual,” and “strategy.”

Our initial intention was to select studies that had been developed or revised in the period 2010–2016, yet we did locate a limited number of policies that were published prior to that period, dating as far back as 1999. In the end, we identified a convenience sample of 15 organizations with existing evaluation policies that had been revised at least once ($N = 32$ policies)² and 20 organizations with a single published evaluation policy. This yielded a total of 52 evaluation policies that serve as the basis for the analysis (see Table 1 for details). The bilateral organizations were headquartered in Australasia ($n = 3$), East Asia ($n = 2$), North America ($n = 3$), and Europe ($n = 7$). Most multilateral organizations belonged to the United Nations family ($n = 15$) and resided in a wide range of geographic locations. Five ($n = 5$) others were in Europe and Latin America. We consider this sample comprehensive but not exhaustive. For more detail, see Al Hudib (2018).

Analysis

The lead author for the current chapter led the qualitative data analysis of the evaluation policies using NVivo (v. 10). Throughout this process, the second author was consulted as issues emerged (e.g., selection decisions, identification of new codes). In some cases joint decisions were made.

We started coding with Trochim’s (2009) eight category scheme but as the coding proceeded, new categories emerged and others needed reconfiguration. The categories were compared for overlap and were refined throughout the process (see Al Hudib, 2018 for elaboration of data quality assurance protocols including the preparation of contextual summaries and the production of an audit trail).

Trochim’s (2009) evaluation policy framework was generated from a theoretical standpoint, as opposed to an empirical one. Our findings identify 10 broad categories, of which seven correspond with Trochim’s categories:

- evaluation goal (aligned with Trochim),

- evaluation definition (emergent category),
- evaluation role (aligned with Trochim),
- evaluation management (aligned with Trochim),
- evaluation norms, standards, criteria, and principles (emergent category),
- evaluation planning and processes (aligned with Trochim but integrates Trochim's original "evaluation participation" category),
- evaluation use (aligned with Trochim),
- meta-evaluation (aligned with Trochim),
- ECB (aligned with Trochim), and
- competencies and ethics (emergent category).

TABLE 1 Achieved Sample of International Development Evaluation Policies⁴

Region	Organization (year)
Africa (<i>n</i> = 2)	ADB: African Development Bank (2007, 2013)
Australasia (<i>N</i> = 3)	AusAID: Australian Aid Agency (2012) ADA: Australian Development Agency (2009) NZAID: New Zealand Aid Program (2014)
North America (<i>n</i> = 5)	CIDA: Canadian International Development Agency (2004*, 2013*) IDRC: International Development Research Centre (Canada) (2013) USAID: United States Agency for International Development (2012, 2016)
East Asia (<i>n</i> = 2)	JICA: Japan International Cooperation Agency (2016) MOFA: Ministry of Foreign Affairs of Japan (2012)
Europe/UK (<i>n</i> = 11)	DANIDA: Danish Development Cooperation (2006, 2012) AECID: Spanish Agency for International Development Cooperation (2013) ADF: French Development Agency (2013) BZ: Dutch Ministry of Foreign Affairs (2009) SDC: Swiss Agency for Development and Cooperation (2013) DFID: UK Department for International Development (2009, 2013, 2014*) SIDA: Swedish International Development Cooperation Agency (1999, 2001**, 2004*)
United Nations (<i>n</i> = 23)	ILO: International Labour Organization (2005, 2012) UNICEF: United Nations International Children's Emergency Fund (2007, 2013) UNIDO: United Nations Industrial Development Organization (2006, 2015) WIPO: World Intellectual Property Organization (2010, 2016) UNESCO: United Nations Educational, Scientific and Cultural Organization (2007, 2015) WHO: World Health Organization (2011, 2012) UNDP: United Nations Development Program (2006, 2016) UNEG: United Nations Evaluation Group (2005*, 2016*) IFAD: International Fund for Agricultural Development (2011) UNITAR: United Nations Institute for Training and Research (2012) UNFPA: United Nations Population Fund (2013) MDG-F: Millennium Development Goals Achievement Fund (2012) OCHA: Office for the Coordination of Humanitarian Affairs (2012) UNEP: United Nations Environment Program (2009) UNHCR: United Nations High Commissioner for Refugees (2010)
International (<i>n</i> = 4)	IFRC: International Federation of Red Cross and Red Crescent Societies (2011) OECD: Organisation for Economic Co-operation and Development – Development Assistance Committee (2010*) IUCN: World Conservation Union (2013, 2015)
Latin America and the Caribbean (<i>n</i> = 1)	ECLAC: Economic Commission for Latin America and the Caribbean (2014)

*Not evaluation policy per se - Evaluation guidelines, manual, strategy, norms, or quality standards.

These dimensions capture the dynamic range and multiple manifestations of organizational evaluation processes. We argue that, taken as a set, they have great potential to influence organizational capacity to do and use evaluation, depending on how they are specified in a given evaluation policy. To what extent do such specifications lend themselves to organizational ECB?

FINDINGS

In this section, we summarize descriptive findings associated with each of the 10 identified categories. At the end of each summary, we comment on the implications of the descriptive findings for the relationship between evaluation policy and organizational ECB. In the text to follow we refer to organizations by their acronyms; full organization names appear in Table 1.

First, we provide a revised version of Trochim's definition of evaluation policy. Based on a close examination of the elements of the evaluation policies included in this study, we decided to expand Trochim's definition (presented in "Evaluation Policy") to make it more inclusive and comprehensive. This study highlights that evaluation policies are concerned with much more than "rules and principles" as they guide decisions and actions that start before and extend beyond the stage of "doing evaluation" as is suggested by Trochim's definition. Policies also include norms and standards that shape and influence the organizational culture. They further include guidelines for supporting and developing organizational and national evaluation capacities. Thus, we offer a revised definition of evaluation policy that captures its actual and practical meaning within an organization:

Evaluation policy is any principle, 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, structures or systems to support and manage ongoing evaluation functions, processes, practices and use.

Evaluation goals

Accountability, learning, and developing evaluative culture are the three major goals of evaluation that emerged in association with this evaluation policy category. Over half of the evaluation policies identified the promotion of accountability as *the* principal goal, often the main reason why the policy was established in the first place. Accountability was often framed in terms of value for money and the demonstration of aid impact. Following is a typical excerpt by way of illustration.

...evaluation is one of several instruments holding the Ministry of Foreign Affairs accountable for the choices and actions taken to meet the overall and specific objectives of Danish development cooperation. (DANIDA, 2012, p. 7)

Thus, evaluation policies' goal of accountability is heavily focused on using evaluation to improve transparency and accuracy in responding to external accountability demands.

Learning as a goal is explicitly identified in almost half of the evaluation policies we looked at and is stated as a principal goal in all more recent evaluation policies, whether newly developed or revised. A focus on learning reflects many organizations' commitment to evaluation as a system to increase knowledge and understanding, which in turn promotes continuous improvement. It also supports the achievement of organizational

goals through enhanced abilities to deal with changing contexts. Consider the following excerpts:

By supporting organizational learning and accountability, evaluation aims to help UNICEF continually to improve its performance and results. (UNICEF, 2013, p. 4)

The purpose of evaluations is learning from experience. With the assistance of evaluations, successes and failures can be interpreted. Based on those experiences, both current and future projects and programmes can be improved. (ADA, 2013, p. 4)

A number of evaluation policies also note that the goals of evaluation include the development of an evaluative culture within the organization. Such goals provide a clear and obvious link to ECB concerns. For example:

The Evaluation Policy identifies three complementary objectives for evaluation: learning, accountability, and promoting an evaluation culture. (AFDB, 2013, p. 6)

In some cases, ECB goals refer not just to the host organization but to country-level partners as well.

Relationship to ECB: There seems to be considerable interest in conceptualizing evaluation as leverage for learning, and complimentary to the accountability function, which is likely to be favorable to ECB (Cousins & Bourgeois, 2014). Yet Kogan (2018) argues that learning and accountability in evaluation policies are often conflated in aid evaluation policies. She asserts that ‘learning’ focused on theories undergirding social change would make evaluations more useful. Regardless, significant attention to the learning function of evaluation holds great promise facilitating organizational ECB.

Evaluation definition

The majority of the evaluation policies (44, or 85%) include a clear definition of the evaluation concept, the second most frequently observed category. Such was previously covered by Trochim (2009) under “goals” but we would argue that its salience warrants explicit identification. Several of the definitions took the added step of differentiating evaluation from other closely related functions such as audit, inspection, and performance measurement. An example follows:

Evaluation is defined as an evidence-based assessment of strategy, policy or programme and project outcomes, by determining their relevance, impact, effectiveness, efficiency and sustainability... It should not be confused with implementation monitoring and reporting, audit, inspection, investigation or assessment of individual performance. (ILO, 2005, p. 2)

In some cases, the definition referred to both learning and accountability goals and foreshadowed intended uses of evaluation.

Relationship to ECB: Given evaluation policy’s use as a communication tool, many organizations acknowledge that definitional clarity is both necessary and important. This policy element is therefore important because it makes explicit the meaning, function,

and use of evaluation within the organization. McDavid and Hughes (2007) previously expressed concern about evaluation losing ground to audit and management consulting professions. It is essential that evaluation stake out its territory and highlight the unique contributions it can make. Definitional clarity, therefore, is likely to be favorable to organizations because it provides exactness about what evaluation is and is not and about what it can be expected to accomplish. In this way, it can provide guidance for individual and organizational ECB goal setting (Labin et al., 2012; Preskill & Boyle, 2008).

Evaluation role

This category focuses on the evaluation responsibilities of various organization stakeholders. Since evaluation is a shared function within the organization, it requires not just cooperation among many different actors but also the distribution of responsibilities and accountabilities among them. We observed explicit reference to evaluation roles and responsibilities in 43 (83%) evaluation policies although the extent of elaboration varied. Here is a typical excerpt:

The Director General is responsible for safeguarding the independence of the Evaluation Section by: (a) Ensuring compliance with the Evaluation Policy set out herein.... (b) Allocating adequate resources—human and financial—to ensure the evaluation function can be carried out professionally.... (WIPO, 2010, p. 6)

Well-articulated roles and responsibilities logically foster greater efficiency and effectiveness in the evaluation function and limit redundancy in workloads, critical process omissions, or other shortcomings.

Many of the evaluation policies we examined even show examples of the intended use of evaluation results being incorporated into the roles and responsibilities of identified members within the organization. Here is an illustration:

IUCN senior management ... are the primary users of M&E results. These results are used to improve performance, perform adaptive management, control risk, enable learning and provide accountability. (IUCN, 2015, p. 12)

Relationship to ECB: Clarity about evaluation roles and responsibilities is likely to be favorable to ECB because it carries potential to build internal cooperation and ownership of the evaluation function, which in turn may contribute to a culture of systemic and evaluative thinking (Buckley et al., 2015; Urban et al., 2021). Buckley et al. (2015) assert that evaluative thinking “is essentially critical thinking applied to evaluation contexts” (p. 375). The development of such a culture is likely to aid in the achievement of organizational goals or mission.

Evaluation management

This category, overtly recognized in 39 (75%) evaluation policies included in the study, focuses on important aspects of the evaluation management function, including evaluation budgeting and resource allocation, evaluation-related human resources management and evaluation timing.

Consider the following illustration:

Each unit that manages program funds and designs and implements projects must conduct at least one evaluation per project. The evaluation must be timed so that the findings will be available as decisions are made about strategies, project designs, and procurements. (USAID, 2016, p. 9)

The majority of the evaluation policies highlight the importance of budgeting when managing the evaluation, most being explicit, some even including guidelines for determining specific allocations:

A dedicated budget line between 3% and 5% should be included for all evaluations of interventions above 200,000 Swiss francs. This approach is consistent with best international practice. (IFRC, 2011, p. 11)

Some evaluation policies articulate guidelines for planning and calculating evaluation costs. Such guidelines are intended to help manage, analyze, and evaluate financial performance.

Relationship to ECB: Sound human resource management and budget planning are integral to an organization's capacity to carry out evaluations effectively. While effective managers are likely to promote the efficient use of resources and productivity, their leadership is most likely to benefit organizational ECB. Administrative leadership and support for evaluation has been shown to underscore the organization's commitment to the evaluation function and carries potential to create a culture of inquiry and organizational learning (Cousins & Bourgeois, 2014; Preskill & Boyle, 2008).

Evaluation norms, standards, principles, and criteria

In our sample, 33 (63%) evaluation policies included such elements as evaluation norms, standards, criteria, and/or principles expected to be considered in planning, designing, implementing, and using evaluation. This is a category that we deemed deserving of explicit identification, whereas such content was covered elsewhere in Trochim's (2009) taxonomy.

The terms *norms* and *standards* are often used interchangeably and refer to how the quality of an evaluation will be judged and can therefore be used for planning purposes. For example, the norms for evaluation of UNDP state that evaluation should be independent, intentional, transparent, ethical, impartial, of high quality, timely, and useful.

Criteria, on the other hand, provide a normative framework used to determine the merit or worth of an intervention upon which evaluative judgments are made. For example, OECD-DAC defined six criteria for development evaluation: relevance, coherence, effectiveness, efficiency, impact, and sustainability. Finally, *principles* most often referred to how criteria are expected to be applied. OECD-DAC³ stipulates "(i) criteria should be applied thoughtfully to support high quality, useful evaluation" and (ii) "the use of criteria depends on the purpose of the evaluation. The criteria should not be applied mechanistically".

Important to note is that most UN organizations make explicit reference to evaluation standards and norms developed by the UNEG and several other multilateral and bilateral agencies reference those, as well as evaluation criteria and principles developed by the OECD-DAC. Such norms and standards and criteria outline professional expectations, ensure consistency throughout the organization's evaluation process, and

ultimately provide evaluators and managers with key information about the quality of the evaluations.

Relationship to ECB: Evaluation policies that explicitly include clear norms and principles of evaluation help to foster an organizational culture of inquiry and evaluative thinking, where members are aware of and agree upon the value of the evaluation function and its role within the organization (Archibald et al., 2016; Cousins & Bourgeois, 2014). Such elements help to operationalize expectations for quality assurance that may be explicit or implicit in evaluation definitions. They also may help decision makers, policy-makers, and other commissioners of evaluation define their expectations for evaluation quality and scope.

Evaluation planning and processes

We renamed this category given an observed strong link to planning as well as implementing evaluation. Comprised within are evaluation planning and design; logic and methods; and participation in the process, including the involvement of partners (formerly a stand-alone category in Trochim's taxonomy). Thirty of the evaluation policies (58%) included this element; many that did not instead referred to companion documents (guidelines, handbooks, etc.)

Observed references to evaluation planning included identifying and documenting the evaluation's scope, design, objectives, questions, methods, requirements, deliverables, schedule setting, and resources. Invariably, evaluation policies promote the concept that evaluation questions and the other key considerations, such as cost, feasibility, and relevance should determine the selection of the evaluation method, rather than the inverse.

Consideration of methods was also tied to expectations for evaluation use. A number of evaluation policies, especially the revised and/or recently established ones, acknowledge that stakeholders' participation in key stages of the evaluation is critical to the creation of a common understanding of purpose and use. Here is an example:

Involvement of and consultation with stakeholders facilitates consensus building and ownership of the findings, conclusions and recommendations; it also heightens the credibility and quality of the evaluation. (UNIDO, 2015, p. 9)

Some evaluation policies extend beyond the stakeholders' involvement to include expectations regarding the conduct of an evaluation in collaboration with partners, particularly for joint evaluations. Evaluation provides opportunities to develop and sustain relationships between organizations; to increase organizations' ability to leverage resources; to build evaluation capacity within the partnership; and to increase accountability. The following excerpt provides a good illustration:

In support of evaluation goals of accountability, learning and performance improvement, UNICEF will maintain active partnerships at all levels with a wide range of institutions, including United Nations agencies, international financial institutions, governments, evaluation associations, non-governmental organizations, foundations and academic institutions. (UNICEF, 2013, p. 17)

Relationship to ECB: Overt attention to the importance of evaluation planning and design is unmistakably important to developing the organizational capacity to implement evaluation. Yet evaluation policies that carry a strong emphasis on stakeholder

engagement and partnership in evaluation are likely to foster the development of organizational capacity to use it (Bourgeois & Cousins, 2008). Cousins and Chouinard's (2012) review and integration of research on participatory and collaborative approaches found strong connections between such approaches and evaluation use, particularly with regard to learning and making program changes. Stakeholder ownership in these types of evaluation tend to be strong.

Evaluation utilization

About two thirds of the evaluation policies provided some focus on the reporting, dissemination and use of evaluation results, under various subheadings. Consistent with our observations about planning and processes, many of the evaluation policies emphasize the importance of identifying and engaging the intended users of the evaluation with the intent of enhancing use. Consider the following illustrative excerpts:

Evaluation is framed in utility: evaluations must have a clear use and respond to the needs of the user, whether management, a program, a donor, or a network of researchers. (IDRC, 2013, p. 1)

...thinking about the use of an evaluation should start at the planning stage rather than when the final report is submitted. The stakeholders of the evaluation should be consulted in the planning stage to ensure that their questions are addressed by the evaluators. (ILO, 2012, p. 45)

The concept of tailored reporting was also common. There was recognition that different groups of intended users have varying needs for the evaluation report. Some evaluation policies included communication strategies to guide the process of reporting the evaluation to diverse audiences. For example:

It is important to plan a communication strategy for the evaluation report to decide who may need information from the evaluation, and the manner in which it should be communicated. It is more likely that users will find evaluations useful if they meet to discuss the main points face-to-face with the evaluators, instead of just receiving a report for comment. (BZ, 2009, p. 6)

We also observed that many evaluation policies underscore the significance of following up on evaluation findings as an essential part of supporting their use, promoting a culture of organizational learning and enhancing accountability for results. Often a management response to the recommendations arising from the report is required, as we can see in this excerpt:

At the conclusion of an evaluation, a follow-up memo is prepared, taking note of DANIDA's position on the conclusions and recommendations as well as identifying which departments are responsible for the agreed follow-up activities. (DANIDA, 2016, p. 38)

Some evaluation policies also include guidelines and strategies for the dissemination of evaluation findings and documents, specifically as an aspect of evaluation use. The intention is to make the evaluation results transparent and accessible to a wide range of relevant stakeholders. UNICEF, for instance, states:

Each evaluation is uploaded into the Global Evaluation Database within 30 days of completion, to permit global sharing and final quality review. (2013, p. 14)

Relationship to ECB: An emphasis on stakeholders' involvement in the evaluation process (planning, interpreting findings, etc.) not only enhances the prospect of enhanced instrumental and conceptual use of findings, but also creates the opportunity for process use. By virtue of internal and broader stakeholder proximity to evaluation, increased learning and ownership are likely to occur (Patton, 2008). Process use can happen at the individual, team and organization level and there is some reason to believe that it has potential to augment use of evaluation results (Cousins & Bourgeois, 2014; Cousins & Chouinard, 2012).

Meta-evaluation

This evaluation policy category focuses on the review and assessment of evaluations (and in some cases evaluators) as part of the organization's evaluation management and quality control system. The meta-evaluation (evaluation of evaluations) category is identified in 23 (44%) evaluation policies of various types of organizations under various headings such as "quality and monitoring," "quality assurance," "quality control," and "quality standards." The following excerpt illustrates:

...evaluation is a tool for quality assurance and quality control, and as such must satisfy its own quality requirements. The quality of any particular evaluation can be assessed in relation to four broad sets of quality standards, namely those of propriety, feasibility, accuracy and utility. (SIDA, 2014, p. 21)

Yet, meta-evaluation also relates to the pursuit of a better understanding of the role played by the evaluation function in the policy cycle and the accumulation of evaluation knowledge that is useful for policy improvement, accountability or edification. Examples include policies related to the periodic assessment of evaluation activities, the coordination of evaluation activities, and the establishment of evaluation clearinghouses for sharing evaluation information. Consider this excerpt:

Improving the quality and usefulness of evaluations is one of the cornerstones of this Policy. To do so, the SGCID ... in consultation with the actors of Spanish Cooperation, will promote the implementation of a quality control system for evaluations, including the establishing of criteria and procedures for conducting meta-evaluations in order to draw lessons and create guidelines for improvement. (SGCID, 2013, p. 16)

As evident in this example, meta-evaluation can carry the intention to promote continuous improvement and use by engaging key stakeholders' evaluation review.

Relationship to ECB: The meta-evaluation category is pivotal to the creation of a supportive organizational setting that influences not only performance management but also organizational learning (Preskill & Boyle, 2008). Intentional learning, not only about the evaluation function, but also about substantive intervention-relevant knowledge is likely to enhance the use of evaluation and therefore favorable to ECB. For example, Cousins and Bourgeois (2014) found that ECB in some organizations had links to processes in place for harvesting knowledge across evaluation products. Our present findings reveal that this practice is part of some organizational evaluation policies.

Evaluation capacity building

This category focuses on the range of efforts that an organization may undertake in order to integrate evaluation practices into its organizational culture. Many organizations acknowledge the importance of strengthening the evaluation capacity of individuals, organizations, communities, and national governments as key to managing evaluation and improving transparency and mutual accountability. ECB was explicitly identified in 17 (33%) evaluation policies in our sample, 9 having a specific section devoted to the concept with the remainder appearing under various sections.

We found many specific examples where ECB content was not just about developing skills for doing evaluation but also about effectively managing, performing and using evaluations and developing an evaluation culture of valuing evidence and evaluative thinking. For example:

DFID is committed to ensuring high quality evaluation. Supporting quality in evaluation means investing in improvements to the process of commissioning and managing evaluations, strengthening staff capacity and capability to commission and manage evaluations, and promoting the uptake of evaluation throughout the organisation and beyond. (DFID, 2013, p. 11)

Many of the agencies included in the sample focus on partner country ECB with aims that include strengthening domestic accountability, facilitating collaboration with partners, and improving development effectiveness at the country level. Consider the following illustration:

Reflecting our commitment to ownership, alignment, harmonisation, mutual accountability and managing for results, we work with our partners on evaluations to increase ownership and use of our evaluations. Partners' involvement also contributes to building their evaluation capacity. (NZAID, 2014, p. 5)

While evaluation capacity is sometimes developed through targeted training or other specific capacity-building programs, a number of evaluation policies emphasize the fact that the evaluation process itself is an opportunity for learning and is a basis on which to leverage and strengthen evaluation capacities.

Relationship to ECB: Obviously, the connection here is quite direct. Naturally, the inclusion of an ECB component in evaluation policies should help to build an evaluation performance culture both internally and within external partner organizations (and national governments). This element can work to support results-based management and policy-making, to strengthen accountability relationships, and to enhance learning. Such may be best accomplished through a balance between direct training and guidance and indirect ECB (i.e., learning by doing) (Chaplowe & Cousins, 2016).

Competencies and ethics

This emergent category focuses on the qualifications, skills, expertise, and ethics that evaluators need in order to fulfill their roles and responsibilities within the evaluation process and thus ensure the credibility and quality of the process. Of the evaluation policies we studied, 15 (29%) included such content as evaluator qualifications, evaluation competencies, and evaluation professionalism and ethics – elements to be considered when recruiting evaluators as well as when planning, conducting, reporting, and using an evaluation.

Evaluation policies that explicitly identify competencies indicate their commitment to high-quality evaluations and to better performance management of the evaluation function in a manner that ensures the integration and practice of the organization's values. Here is an illustration:

The Evaluation Group issues formal job descriptions and selection criteria that state the necessary professional requirements, skills and competences for hiring evaluation managers and evaluators, as well as contracting evaluation experts. (UNIDO, 2015, p. 9)

Some evaluation policies even include a model for evaluation competencies in order to ensure a proper implementation of the evaluation policy, manage professional standards and strengthen evaluation quality. The DFID evaluation policy provides a good example:

The evaluation accreditation model and evaluation competencies are designed to ensure high professional standards are maintained. (2013, p. 11)

Our analysis reveals an overlap between competencies and ethical principles when it comes to practicing evaluation. The analysis shows that most of the policies focus on factors that evaluators and evaluation managers must respect at all stages of the evaluation process, namely human rights and gender and cultural diversity, especially with the many calls for a more active role for stakeholders in all aspects of the evaluation process. For instance:

Stakeholder rights and interests shall be respected, particularly with regard to respecting culture and customs, fundamental values, the right to or not to participate, and ensuring confidentiality of individual stakeholders. (IUCN, 2015, p. 10)

Relationship to ECB: It can be argued that success in evaluation hinges on recruiting and developing professionally competent evaluators, who embrace the organization's values and ethical principles. By clearly communicating ethical principles within and through the evaluation policy, organizations leave little room for doubt, confusion, or contradiction. This category is likely to be favorable to ECB because it sets the standards for professional and ethical evaluation practices and makes these practices not only part of the evaluation function but also part of the organizational culture (Labin et al., 2012; Preskill & Boyle, 2008).

DISCUSSION

One key outcome of this exploratory study was a revision of Trochim's (2009) definition of evaluation policies. Another was the refinement and extension of his eight-category evaluation policy framework. Through the analysis process, we were able to identify 10 broad categories that capture the goals, processes, and consequences of evaluation policy. A third outcome is that, by virtue of our positive application of the framework, we successfully met Cooksy et al.'s (2009) empirical validation challenge. Yet, perhaps the most significant contribution of the study is the preliminary connections made between evaluation policy and ECB. Of particular salience are the inter-related themes of evaluation use, evaluation purposes, and the critical importance of leadership. Before considering these, however, we address some limitations in our study.

This descriptive, exploratory foray into the connection between evaluation policy and ECB has considerable value, particularly in laying the groundwork for future inquiry. Yet the design is quite limited in permitting causal assertions about the relationship. For example, of high interest to organization and evaluation practitioners would be the relative weight one might ascribe to specific evaluation policy elements. To what extent are some elements more likely to produce desired consequences for organizational evaluation capacity? We can only speculate about this and part of our answer is addressed below as we look at purposes for evaluation. Specifically, our view is that investing in the learning function of evaluation holds great promise for ECB.

A second concern links back to our distinctions between the espoused policy and policy-in-use. It may be fair to say that most organization members would recognize that a considerable gap between the two is not uncommon. To what extent will these policy elements be faithfully implemented in practice? Obviously, their effect on evaluation practice and organizational evaluation capacity will depend quite directly on how well these policy elements are implemented. Given the limitations of our present data, we are unable to comment on this issue. We now turn to a discussion of inter-related themes, beginning with evaluation use.

Data use leads to data valuing is an assertion arising from Cousins and Bourgeois' (2014) multiple case study of organizational evaluation capacity. It implies that by virtue of a combination of pressure and support, if we can get organization members to experience the successful use of evaluation, they would be more likely to embrace and integrate it into their organizational culture. Specifically, they would be more predisposed to the promotion of systems and evaluative thinking and integrating evaluation as an organizational learning system (Archibald et al., 2016; Buckley et al., 2015; Nelson et al., 2021; Urban et al., 2021). Evaluation policy elements that may be conducive to such growth include a focus on learning and conceptual use of findings, as well as the promotion of program stakeholder involvement in evaluation to enhance learning and increase ownership (Cousins & Bourgeois, 2014; Cousins & Chouinard, 2012; Patton, 2008). Furthermore, evaluation policy allowances for participatory and collaborative approaches that implicate stakeholders in planning, design, and implementation activities may also contribute (Cousins, 2020). The latter observation is consistent with recent studies that underscore the advantages of participatory approaches to evaluation in addressing the complex dynamics of programs and in increasing learning from the evaluation (Chouinard, 2013; Lennie & Tacchi, 2014).

The present research demonstrates that evaluation is seen as a mechanism for the promotion of both accountability and learning as fundamental purposes for evaluation, although capacity building was also mentioned with some frequency. In fact, the tension between accountability and learning as justifications for, and results of, evaluation has been a subject of increasing interest (Chouinard, 2013; Feinstein, 2012; Guijt, 2010). While both functions were cited in most evaluation policies in our sample, a challenge may be that their meaning is largely assumed rather than being explained (Kogan, 2018). Is there a clear definition of learning? In the absence of a clear definition, the risk is that learning becomes a mere by-product of accountability. Organizational evaluation policies would do well to reconsider learning and accountability concepts in the interest of ECB. Hay (2010), for example, offers that more emphasis could be put on 'downward accountability,' or accountability to aid recipients and intended beneficiaries, than is presently the case. Kogan (2018) advocates a focus on exploring and validating program theory. Both ideas align well with the capacity building function. Also, "internal accountability" might be reframed as processes of organizational learning and change that would contribute to internal learning and decision making. Evaluation policy focus could be placed on use-oriented strategies (e.g., utilization-focused evaluation, Patton, 2008), which would be responsive to the learning needs of the organization members (Abma, 2005).

Five of the 10 evaluation policy categories identified and used in this study – evaluation role, evaluation management, evaluation use, ECB, and competencies and ethics – speak directly to the important role of organization managers in fostering evaluation capacity. The various skills and areas of expertise required to attend to these categories of the evaluation policy clearly indicate the fundamental role of organization leaders in establishing, building, and sustaining organizational capacity for evaluation. Leaders can do this by: (i) understanding how strategy and evaluation are (or should be) interconnected, (ii) providing adequate resources for evaluation, (iii) being active consumers of evaluation information, and (iv) using evaluation as a means for ongoing organizational learning (Cousins & Bourgeois, 2014; Preskill & Torres, 1999). The benefits of supportive and committed leadership to facilitate and manage the implementation of ECB efforts and to internalize norms and process ownership should not be underestimated. Yet, Labin et al. (2012) observed in their review of ECB theory and practice, “Leadership was the least frequently targeted organizational factor and the least frequently reported organizational outcome” (p. 321). Evaluation policy, we would argue, has a role to play here.

Ongoing inquiry

Our own research program continues to make some positive strides in this domain of inquiry. In a study based on the findings of the present one, we interviewed a range of evaluation policy and ECB scholars and contributors. Emerging was an ecological conceptual framework to guide thinking about the role of evaluation policy in capacity building (Al Hudib & Cousins, 2021). Subsequent validation of the framework in a cross-cultural context was encouraging (Al Hudib & Cousins, 2020). We offer this framework as an evidence-informed guide to research and inquiry on the ECB/evaluation policy interface. In the present study, we included existing and revised evaluation policies from 15 organizations, but our focus was restricted to descriptive content analysis at the policy level. Currently underway is a mixed method, longitudinal study in which we are systematically examining evaluation policy changes within organizations to determine the extent to which they are conducive to fostering ECB.

We recommend that ongoing research consider alternative methodologies, perhaps including quantitative or mixed methods approaches and longitudinal elements. It would also be interesting to examine the evaluation policy revision process from a case study perspective. To what extent are ECB principles and considerations taken into account within evaluation policy reform activities? Finally, much of what we have learned about the interface between evaluation policy and organizational ECB comes from the perspective of evaluation practitioners or scholars. Future studies could investigate how the framework resonates with evaluation users, for example, organization or program managers and higher-level decision and policy makers. Such studies could also focus on one or more particular types of organization (e.g., non-profit organizations) in order to generate insights related to issues concerning ECB and the major variables moderating the role of evaluation policy in those environments.

A word about practice

We are averse to providing practical advice to organization members and evaluators given the exploratory nature of our study. Yet, tentatively, we might glean some implications for practice from the foregoing discussion. First, practitioners would do well to take necessary steps to reduce the gap between the written evaluation policy and the informally

understood policy. Second, consider in ECB not just the organizational capacity to do evaluation but also the capacity to use it. This might imply privileging the learning function of evaluation to the extent possible. While embracing the need for accountability, take steps to ensure that it is genuine and does not become a symbolic, checklist activity. Finally, if organizations truly embrace the goal of enhancing organizational evaluation capacity, leadership is essential. There is a symbolic role for organization leaders to engage with evaluation, and to promote it. Perhaps most important, administrative leaders should take every opportunity to demonstrate the power and value that evaluation can bring to the organization.

ACKNOWLEDGEMENT

The research reported in this chapter is based on the first of three studies from the lead author's doctoral dissertation at the University of Ottawa (Al Hudib, 2018). The authors are grateful to the members of the thesis examining committee for their input and guidance.

ORCID

Hind Al Hudib  <https://orcid.org/0000-0003-2514-1945>

J. Bradley Cousins  <https://orcid.org/0000-0002-0684-6766>

END NOTES

¹ Evaluation 'process use' connotes learning and evaluation skill development by virtue of proximity or exposure to evaluation implementation (Patton, 2008).

² An analysis of evaluation policy change is beyond the scope of the present descriptive study. See Discussion section for commentary on this prospect.

³ Retrieved from <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm#:~:text=The%20criteria%20should%20be%20applied,evaluated%2C%20and%20the%20stakeholders%20involved.>; November 23, 2019

⁴ The sample was drawn in 2016 as part of Al Hudib's doctoral research (2018). Many of the organizations included in the sample have since revised their evaluation policies. Documents used for this analysis are available on request.

REFERENCES

- Abma, T. A. (2005). Responsive evaluation: Its meaning and special contribution to health promotion. *Evaluation and Program Planning*, 28, 279–289. doi: 10.1016/j.evalprogplan.2005.04.003
- Al Hudib, H. (2018). *The role of evaluation policy in organizational capacity to do and use evaluation*. Unpublished Ph.D. dissertation, Ottawa: University of Ottawa.
- Al Hudib, H., & Cousins, J. B. (2020). Evaluation policy and organizational evaluation capacity building: Application of an ecological framework across cultural contexts. *Journal of MultiDisciplinary Evaluation*, 16(36), 37–56.
- Al Hudib, H., & Cousins, J. B. (2021). Understanding evaluation policy and organizational capacity for evaluation: An interview study. *American Journal of Evaluation*, Online first.
- Archibald, T., Sharrock, G., Buckley, J., & Cook, N. (2016). Assumptions, conjectures, and other miracles: The application of evaluative thinking to theory of change models in community development. *Evaluation and Program Planning*, 59, 119–127. doi: <https://doi.org/10.1016/j.evalprogplan.2016.05.015>
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. <https://doi.org/10.1177/1098214015581706>
- Bourgeois, I., & Cousins, J. B. (2008). Informing evaluation capacity building through profiling organizational capacity for evaluation: An empirical examination of four Canadian Federal Government organizations. *Canadian Journal of Program Evaluation*, 23(3), 127–146.
- Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*, 34(3), 299–319. <https://doi.org/10.1177/1098214013477235>
- Bourgeois, I., Toews, E., Whynot, J., & Lamarche, M. K. (2013). Measuring organizational evaluation capacity in the Canadian Federal Government. *Canadian Journal of Program Evaluation*, 28(2), 1–19.
- Bourgeois, I., Hotte, N., Simmons, L., & Osseni, R. (2016). Measuring evaluation capacity in Ontario public health units. *Canadian Journal of Program Evaluation*, 31(2), 165–183. <https://doi.org/10.3138/cjpe.306>

- Buetti, D., Bourgeois, I., & Savard, S. (2019). Modélisation des capacités organisationnelles en évaluation dans le secteur communautaire et implications pour le contexte québécois. *ANSERJ*, 10(2), 16–39. <https://doi.org/10.29173/anserj.2019v10n2a287>
- Chaplowe, S. G., & Cousins, J. B. (2016). *Monitoring and Evaluation Training: A systematic approach*. Sage.
- Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*, 34(2), 237–253. <https://doi.org/10.1177/1098214013478142>
- Cheng, S.-H., & King, J. A. (2017). Exploring organizational evaluation capacity and evaluation capacity building: A Delphi study of Taiwanese elementary and junior high schools. *American Journal of Evaluation*, 38, 521–539. <https://doi.org/10.1177/1098214016672344>
- Compton, D., Baizerman, M., & Stockdill, S. H. (Eds.). (2002). *The art, craft and science of evaluation capacity building: New directions for evaluation*. No 93. Jossey Bass. <https://doi.org/10.1002/ev.38>
- Cooksy, L. J., Mark, M. M., & Trochim, W. M. K. (2009). Evaluation policy and evaluation practice: Where do we go from here? In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.). *Evaluation policy and evaluation practice: New directions for evaluation* (Vol. 123, pp. 103–109). Wiley. <https://doi.org/10.1002/ev.308>
- Cousins, J. B. (Ed.). (2020). *Collaborative approaches to evaluation: Principles in use*. SAGE.
- Cousins, J. B., Amo, C., Bourgeois, I., Chouinard, J., Goh, S., & Lahey, R. (2008). Organizational capacity to do and use evaluation: Results of a pan Canadian survey. *Canadian Journal of Program Evaluation*, 23(3), 1–25.
- Cousins, J. B., & Bourgeois, I. (Eds.). (2014). *Organizational capacity to do and use evaluation. New directions for evaluation*. No. 141. Jossey Bass.
- Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: A review and integration of the research base*. Information Age Press.
- Feinstein, O. N. (2012). Evaluation as a learning tool. In S. Kushner & E. Rotondo (Eds.), *Evaluation voices from Latin America: New directions for evaluation* (Vol. 134, pp. 103–112). Wiley. <https://doi.org/10.1002/ev.20022>
- Guijt, I. (2010). Exploding the myth of incompatibility between accountability and learning. *Capacity Development in Practice* (pp. 277–291). EarthScan.
- Hay, K. (2010). Evaluation field building in South Asia: Reflections, anecdotes, and questions. *American Journal of Evaluation*, 31(2), 222–231. doi: [10.1177/1098214010366175](https://doi.org/10.1177/1098214010366175)
- Kogan, L. (2018). What have we learned here? Questioning accountability in aid policy and practice. *Evaluation*, 24(1), 98–112. <https://doi.org/10.1177/1356389017750195>
- Labin, S., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33, 307–338. <https://doi.org/10.1177/1098214011434608>
- Leeuw, F. L. (2009). Evaluation policy in the Netherlands. In W.M.K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.). *Evaluation policy and evaluation practice: New directions for evaluation* (Vol. 123, pp. 87–102). Wiley. <https://doi.org/10.1002/ev.307>
- Lennie, J., & Tacchi, J. (2014). Bridging the divide between upward accountability and learning-based approaches to development evaluation Strategies for an enabling environment. *Evaluation Journal of Australasia*, 14(1), 12–23.
- Mark, M. M., Cooksy, L. J., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.). *Evaluation policy and evaluation practice. New directions for evaluation* (Vol. 123, pp. 3–11). Wiley. <https://doi.org/10.1002/ev.302>
- McDavid, J., & Huse, I. (2007). Will evaluation prosper in the future? *Canadian Journal of Program Evaluation*, 21(3), 47–72.
- Nacarrella, L., Pirkis, J., Kohn, F., Morley, B., Burgess, P., & Blashki, G. (2007). Building evaluation capacity: Definitional and practical implications from an Australian case study. *Evaluation and Program Planning*, 30(3), 231–236. <https://doi.org/10.1016/j.evalprogplan.2007.05.001>
- Nelson, A. G., King, J. A., Lawrenz, F., Reich, C., Bequette, M., Pattison, S., Kollmann, E. K., Illes, M., Cohn, S., Iacovelli, S., Cardiel, C. L. B., Ostgaard, G., Goss, J., Beyer, M., Causey, L., Sinkey, A., & Francisco, M. (2021). Using a complex adaptive systems perspective to illuminate the concept of evaluation capacity building in a network. *American Journal of Evaluation*, 2019(2), 214–230. <https://doi.org/10.1177/1098214018773877>
- Nielsen, S. B., Lemire, S., & Skov, M. (2011). Measuring evaluation capacity: Results and implications of a Danish study. *American Journal of Evaluation*, 32(3), 324–344. <https://doi.org/10.1177/1098214010396075>
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459. <https://doi.org/10.1177/1098214013499439>
- Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. Sage.
- Stockdill, S. H., Baizerman, M., & Compton, D. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. In D. Compton, M. Baizerman, & S. H. Stockdill (Eds.), *The art, craft and science of evaluation capacity building: New directions for evaluation*, No. 93 (pp. 7–25). Jossey-Bass. <https://doi.org/10.1002/ev.39>
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system*:

Lessons from cooperative extension. New Directions for Evaluation (Vol. 120, pp. 55–69). Wiley Periodicals. <https://doi.org/10.1002/ev.276>

Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. In W. M. K. Trochim, M. M. Mark, & L. J. Cooks (Eds.), *Evaluation policy and evaluation practice. New directions for evaluation* (Vol. 123, pp. 13–32). Wiley. <https://doi.org/10.1002/ev.303>

Urban, J. B., Archibald, T., Hargreaves, M., Buckley, J., Hebbard, C., Linver, M. R., & Trochim, W. M. (2021). Introduction to relational systems evaluation. In J. B. Urban, T. Archibald, M. Hargreaves, J. Buckley, C. Hebbard, M. R. Linver, & W. M. Trochim (Eds.), *Relational systems evaluation: New directions for evaluation* (Vol. 169, pp. 11–18). Wiley. <https://doi.org/10.1002/ev.20444>

AUTHOR BIOGRAPHIES

Hind Al Hudib, PhD is an independent researcher who recently completed a postdoctoral fellowship at the Centre for Research on Educational and Community Services (CRECS), University of Ottawa. Her interests include evaluation policy and organizational capacity for evaluation, and collaborative approaches to evaluation.

J. Bradley Cousins, PhD is Professor Emeritus and former Director of the Centre for Research on Educational and Community Services, University of Ottawa. He is also Distinguished Visiting Professor at Claremont Graduate University. His primary research interests include collaborative approaches to evaluation, evaluation use, and evaluation capacity building.

How to cite this article: Al Hudib, H., & Cousins, J. B. (2022). Evaluation policy and organizational evaluation capacity building: a study of international aid agency evaluation policies. *New Directions for Evaluation*, 2022, 29–48. <https://doi.org/10.1002/ev.20494>