

The Role of Evaluation Policy in Organizational Capacity to Do and Use Evaluation

by

Hind Al Hudib

Faculty of Education

Faculty of Graduate and Postdoctoral Studies

University of Ottawa

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the Ph.D.

© Hind Al Hudib, Ottawa, Canada, 2018

ABSTRACT

Despite the recent calls made by scholars in the evaluation field regarding the importance of evaluation policy and its influence on evaluation practice, there remains a lack of empirical evidence pertaining to the relationship between evaluation policy and evaluation capacity building (ECB). This study sought to explore the role of evaluation policy in building, or in impeding, organizational capacity to do and use evaluation. Through three interconnected studies—a review of an extensive sample of evaluation policies; interviews with scholars and practitioners from Canada, the United States, and Europe; and focus groups held with evaluation community members in Jordan and Turkey—the research identified a set of 10 categories of evaluation policy and proceeded to develop and validate an ecological framework depicting the relationship between evaluation policy and organizational capacity to do and use evaluation. The findings suggest that the role of evaluation policy in building organizational capacity for evaluation is moderated by a number of variables operating at the contextual, organizational and individual levels and that an in-depth understanding of the dynamic, unfolding and ongoing connections between ECB, on the one hand, and the broader social, economic, political and cultural systems associated with an organization, on the other, is essential in focusing ECB efforts. While the findings reveal that the role of evaluation policy in leveraging organizational evaluation capacity has been limited, they also show some evidence that if an evaluation policy is carefully designed to privilege learning as a central and desirable function of evaluation it will be more likely to have a positive influence on the organizational capacity to do and use evaluation. The investigation helps to advance understanding of these connections and provides some insight into the components of evaluation policies and the role that they might play in shaping the future of evaluation practice. This thesis makes an important contribution to the body of knowledge on organizational evaluation capacity. Although much has been published in the evaluation literature on ECB, its relationship to evaluation policy has not been explored or described based on empirical data. The main practical implication of the research is the possibility for organizations seeking to develop evaluation policies that are ECB-oriented to use the ecological framework and the set of evaluation policy categories as guides. Similarly, organizations that are seeking to review and update their current policies to make them more ECB-friendly stand to benefit in this way. Future research may focus on expanding the scope of the framework and its applicability for different types of organizations in different contexts. Finally, it is argued that the development of policies designed to promote learning is a necessary step towards the advancement of evaluation practice.

DEDICATION

This thesis is dedicated first and foremost to my husband, Ahmed. His remarkable patience and unwavering love, encouragement and continuous support were invaluable throughout all of my graduate studies.

I also dedicate this thesis to my daughters, Shomok, Ghada and Sara. They provided the inspiration necessary for me to complete this process. I hope to be a good role model to them, now and in the years to come.

ACKNOWLEDGMENTS

The development of this Ph.D. would not have been possible without the support and guidance of the many people who accompanied me during this process. I first would like to express my deep and sincere gratitude to my thesis supervisor, Dr. Brad Cousins, for his continuous support and encouragement throughout my doctoral studies. I consider myself to be truly fortunate to have had the possibility to work closely with him over the last few years. I appreciate all of the opportunities that he has provided for expanding my understanding of evaluation and of evaluation capacity building. His availability and his generosity with his time, especially during the thesis writing, made it possible for me to keep my research on track.

I would like to thank the members of the thesis committee, Dr. Ruth Kane, Dr. Peter Milley, Dr. Richard Maclure, Dr. Katherine Moreau and my external examiner Dr. Melvin Mark for their valuable guidance and inspiring discussion. My thesis is much better because of their feedback.

I also would like to thank all of the interviews participants who took time out of their busy lives to share their experiences and reflections with me. I could never have written this thesis without their candour, interest and participation.

I further want to acknowledge all of the people who generously participated in my research in both Turkey and Jordan. I sincerely appreciate the openness with which they shared their views and experiences with me.

Finally, I would like to thank my friends and family for their unfailing support and encouragement. This accomplishment would not have been possible without them.

TABLE OF CONTENTS

ABSTRACT	ii
DEDICATION.....	iii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	ix
LIST OF TABLES	x
Chapter 1: Introduction	1
Statement of the Problem	1
Background	1
Guiding Research Questions	5
Contributions to the Field of Evaluation	5
Context of the Study	6
Overview of Methodology and Research Plan.....	7
Overview of the Thesis.....	9
Chapter 2: Review of the Scholarship on Evaluation Policy and Capacity Building	10
What is Evaluation?.....	10
Evaluation Capacity Building.....	11
Profile of Growth in Organizational Evaluation Capacity.....	13
Capacity to Do Evaluation.....	14
Capacity to Use Evaluation.....	15
Understanding Policy.....	17
Policy Formulation and Implementation	18
Policy and Power	21
Espoused Policy and Policy in Use.....	24
What is Evaluation Policy?.....	25
Why Does Evaluation Policy Matter?.....	26
Evaluation Policy as Part of the Evaluation System	27
Categories of Evaluation Policy.....	28
Conceptual Framework.....	29

Chapter 3: Methodology	32
Overall Research Design	32
Epistemological Orientation	36
Study 1: Evaluation Policy Analysis (Research Question 1)	37
Sample Selection for Study 1	38
Study 1 Sample Characteristics	38
Study 1 Review Strategy and Analysis	42
Study 2: Interviews with Evaluation Scholars and Practitioners (Research Questions 1, 2, & 3)	44
Study 2 Sample Selection and Characteristics	44
Study 2 Data Collection	45
Study 2 Data Analysis	46
Study 3: Focus Groups with MENA Evaluation Community Members	47
Study 3 Sample Selection and Characteristics	48
Focus group 1: The Turkish Monitoring and Evaluation Society (TMES)	48
Focus group 2: Jordanian Development Evaluation Association (EvalJordan)	49
Study 3 Data Collection	50
Study 3 Data Analysis	51
Ensuring Data Quality	51
Chapter 4: Evaluation Policy Analysis Findings	54
Comprehensive Review and Analysis of Evaluation Policies	54
Evaluation Goal	55
Evaluation Definition	58
Evaluation Role	59
Evaluation Management	62
Evaluation Norms and Standards	64
Evaluation Planning and Processes	65
Evaluation Utilization	68
Meta-Evaluation	72
Evaluation Capacity Building	74
Competencies and Ethics	77
Other Findings	80
A New Definition of Evaluation Policy	83

Summary.....	84
Chapter 5: Findings from Interviews with Evaluation Scholars and Practitioners	85
Major Findings.....	86
Macro Level: Socio-political, Historical, Cultural and Economic Context	88
Meso Level: Organizational Context	93
Capacity to do evaluation.....	93
Capacity to use evaluation	99
Micro level: Evaluation Stakeholders.....	107
Capacity to do evaluation.....	108
Capacity to use evaluation.	110
Observations	114
Summary.....	115
Chapter 6: Findings from the Focus Group with MENA Evaluation Community	
Members	117
Major Findings.....	117
Turkish Focus Group	118
Macro level: Social, political, historical, and economic context.	118
Meso level: Organizational context.	121
Micro level: Evaluation stakeholders.....	122
Jordanian Focus Group	123
Macro level: Social, political, historical, and economic context.	123
Meso level: Organizational context.	125
Micro level: Evaluation stakeholders.....	127
Summary.....	129
Chapter 7: Discussion and Conclusion	131
Summary of Findings	131
Relevant Themes Emerging from the Literature.....	135
Evaluation Context	135
Evaluation Use.....	138
Evaluation Purpose: Accountability and Learning	139
Organization Leaders' Evaluation Capacity	143
Evaluation Policy as a Capacity Building System.....	145

Limitations of the Study	147
Implications for Practice	149
Implications for Research	151
Theoretical Investigations.....	151
Alternative Methods	151
Different Points of View.....	152
More Exploration.....	152
Concluding Remarks	152
References	155
Appendices.....	168
Appendix A: Dimensions and Sub-dimensions of Organizational-Evaluation Capacity	168
Appendix B: List Evaluation Policies.....	169
Appendix C: Evaluation Policy Data Elaboration	173
Appendix D: Interview Questions	200
Appendix E: Introductory Email (Study 2).....	202
Appendix F: Official Invitation (Study 2).....	203
Appendix G: Informed Consent	204
Appendix H: Research Background (Study 3).....	206

LIST OF FIGURES

Figure 1. Conceptual framework describing the relationships between evaluation policy and organizational capacity to do and use evaluation.	29
Figure 2. Summary of methodology and logical integration of studies.	32
Figure 3. Conceptual basis of research plan.	35
Figure 4. Moderating variables that explain the relationship between evaluation policy and organizational capacity for evaluation.	86

LIST OF TABLES

Table 1 Evaluation Policies (N=52).....	39
Table 2 Distribution of Evaluation Policies by Geographical Region.....	42

Chapter 1: Introduction

Statement of the Problem

There is a lack of empirical knowledge about the role played by evaluation policy in developing organizational capacity for evaluation. In fact, most of the available literature on evaluation capacity building (ECB) focuses on characterizing ECB processes and tries to describe, define, and measure them. However, as some researchers have acknowledged, there remain major deficiencies in our understanding of organizational capacity to do and use evaluation (Cousins et al., 2008; Nacarrella et al., 2007; Nielsen, Lemire & Skov, 2011). Further, currently little if anything is known about the role that evaluation policy plays in organizational capacity for evaluation. Searches of the evaluation literature for studies on evaluation policy for examples of the components of organizational evaluation policy, for information about the influence of evaluation policy on ECB practices, and for context-specific factors and conditions that may facilitate or impede the role of evaluation policy in the integration of evaluation practices within organizations return very few conceptual papers and, to my knowledge, no empirical studies. If we are to gain a deep understanding of organizational capacity to do and use evaluation, it is imperative that we conduct empirical studies on the impact of evaluation policy. This thesis explores these links and addresses these deficiencies in our knowledge by identifying the key variables that moderate the relationship between evaluation policy and organizational capacity for evaluation.

Background

During a period of public budget constraints and an increasingly competitive global economy, the level of international interest in ECB continues to grow among organizations, funders, and evaluators. Evaluators and evaluation scholars who are concerned with building evaluation culture and practices in organizations have been adding to the growing base of theoretical and empirical ECB literature. Many studies have theorized that the enhancement of organizational evaluation capacity, which can be thought of as the competencies and structures required to conduct high-quality evaluation studies (the capacity to do evaluation), as well as the organization's ability to integrate evaluation findings into its decision-making processes (the capacity to use evaluation), is key to the efficient production of quality

evaluations that enable organizations to meet their increasing accountability requirements and to use evaluation for organizational development and decision making (e.g., Bourgeois, Toews, Whynot, & Lamarche, 2013; Boyle, Lemaire, & Rist, 1999; Cousins, Goh, Clark, & Lee, 2004; Sanders, 2002; Stockdill, Baizerman, & Compton, 2002). Light (2004) argues that capacity building involves a range of activities within an organization such as strategic planning, technology upgrades, and management training. The key, he argues, is the link between these activities and organization performance effectiveness.

In recent years, the increasing demand for quality evaluation studies has led many governmental, para-governmental and not-for-profit organizations and agencies to establish evaluation policies to meet their accountability requirements and guide their evaluation practices. For example, in the United States, the 1997 Government Performance and Results Act (GPRA) aims to enhance the efficiency and effectiveness of federally funded programs by compelling agencies to focus on program results, quality of service and customer satisfaction (Radin, 1998). In 2010, the GPRA was updated through the GPRA Modernization Act, which requires the development of strategic plans that specify measurable goals (Office of Management and Budget, 2010). The updated statute reflects the shift in the focus of accountability from the production of reports to the use of performance data to evaluate outcomes.

Similarly, in Canada, a number of federal accountability measures have created an environment in which government agencies are required to set performance indicators and measure outcomes. For example, all government departments are subject to the evaluation requirements set out in the Policy on Results (Government of Canada, 2016) and to government expectations for the implementation of the Results-based Management Framework facilitated through the Secretary of the Treasury Board to undertake evaluation of, monitor and report on program performance.

This trend is by no means restricted to North America. Some international multilateral organizations, such as the United Nations, have recognized the importance of establishing an evaluation policy “to increase transparency, coherence and efficiency in generating and using evaluative knowledge for organizational learning and effective management for results, and to support accountability” (UNDP, 2011).

Evaluation policy can specify a range of requirements related to evaluation. Trochim (2009) notes that 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 deal with how evaluation should be done, what resources are needed, who is responsible for which tasks, what types of expertise are required, what the expectations are for evaluation, and how evaluations should be conducted (Mark, Coosky, & Trochim, 2009). Most of these questions align with considerations for the practice of evaluation as opposed to its use. Recent work on ECB (Cousins & Bourgeois, 2014) has heightened the importance of understanding not just the organizational capacity to *do* evaluation but also the capacity to *use* it. Our knowledge of the latter is comparatively limited.

Although considerable progress has been made in gaining an understanding of what constitutes effective ECB and in identifying its role as a continuous learning and development system, there remain significant theoretical and empirical questions about the impact of evaluation policy on evaluation practice (Mark et al., 2009) and about the organizational capacity to do and use evaluations. Bearing this in mind, practitioners and scholars of evaluation have recently called attention to the importance of policy and its impact on practice (Trochim, 2009). In his role as president of the American Evaluation Association (AEA) Trochim (2008) provided a thoughtful overview for consideration by framing the theme of “Evaluation Policy and Evaluation Practice” for the Association’s annual meeting. In 2009, in the *New Directions for Evaluation* special issue on the same theme, Trochim laid out a theoretical framework for evaluation policy and several ways to further our knowledge about it. Trochim’s emphasis on theorizing evaluation policy reflects a shift in focus from customary evaluation routines to the effect that they have on evaluation practice. However, despite his efforts to keep the discussion going, a review of the relevant literature shows that evaluation policy does not seem to be part of any discourse on how to enhance organizational capacity for evaluation.

Thus far, the extent to which evaluation policy might influence organizational capacity building has not been the focus of any empirical research. In fact, the actual categories of evaluation policy as they might be observed in organizations' policy documents have not yet been empirically identified in the literature. Knowledge of these categories might help organizations identify what areas to target and what efforts are required in order to develop their capacity or attain a higher level of evaluation capacity. This thesis identifies the key categories of evaluation policy as they appear in organizations' policies. The relationship between the evaluation policies and ECB is described through a framework based on the ecological system theory developed by Bronfenbrenner (1979, 2005).

The overarching objective of the present research is to address this gap by drawing attention to the importance of evaluation policy and its influence on organizational capacity for evaluation. Given the fact that ECB has an explicit goal of developing organizational capacity for evaluation, it is critical that we expand and deepen our knowledge about the process by situating it within a broader context of evaluation policy. This study considers research on evaluation in this area, informed by two conceptual frameworks, with attention to how they might be used together to further understanding and be useful to research about ECB. The first is a framework for organizational evaluation capacity developed by Bourgeois and Cousins in 2008 (most recently published in its entirety in 2013); the second is a theoretical framework developed by Trochim (2009) that describes the dimensions of evaluation policy. The new approach offered in this study is an integration of these frameworks that seeks to identify the policy and program levers that can be important at different stages of the ECB process.

Through a multi-study approach, this study sheds light on evaluation policy's influence on the organizational capacity to do and use evaluation in terms of how organizational ECB processes are shaped, reinforced, and accelerated (or impeded) by evaluation policies.

Guiding Research Questions

This research addresses the gaps identified above by examining the following questions:

1. What is the role of evaluation policy in the organizational capacity to do and use evaluation? To what extent is the development of organizational capacity to do and to use evaluation espoused in existing organizational evaluation policies?
2. How does evaluation policy enhance (or impede) the integration of evaluation into organizational culture?
3. What are the contextual facilitators and barriers that influence the role of evaluation policy in enhancing the integration of evaluation into organizational culture?

Since much is already known about the processes of ECB, it is worthwhile to move forward by exploring the influence of evaluation policy and what it might look like on a practical level in relation to organizational capacity for evaluation.

Contributions to the Field of Evaluation

Many evaluation scholars agree that the link between evaluation practice and theory is an area of much-needed inquiry (Cousins & Earl, 1999; Scriven, 1991; Shadish, Cook & Leviton, 1991; Smith, 1993; Worthen, 1990, 2001). Indeed, “We need to add to the empirical knowledge base through carefully developed and executed studies that have the potential to extend our theories and guide our practices, studies that manifestly strengthen the link between theory and practice” (Cousins & Earl, 1999, p. 316).

In addressing this need, this study recognizes that organizational capacity to perform and use evaluation is largely dependent on the actions that organizations take to integrate evaluation practices into their systems. Underpinning this integration is their policy on evaluation. By combining the ECB and policy frameworks mentioned above this study brings a new and important lens to the consideration of evaluation policy and its practical impact. Bourgeois and Cousins’ framework is based on empirical research conducted on Canadian federal government organizations. It would be of considerable value to expand the scope of application of the framework to other types of organizations and to examine its applicability

and relevance in the context of developing countries. By examining the role of policy in organizational evaluation capacity, this research study will add new insights to the contextual factors that facilitate or restrain ECB efforts.

Context of the Study

Organizational evaluation capacity is a moving target. The rapidly changing and evolving character of the world—as evidenced by government reforms, shifting economic conditions, financial crises, major movements in energy prices, climate change, and degradation and depletion of natural resources—has led many governmental, para-governmental and not-for-profit organizations and agencies to develop evaluation policies and strategies to manage their resources and evaluate the impact of their work (Mayne, 2007, 2009; Rist, 2006). As part and parcel of a new public management (NPM) governance regime, leading organizations around the world are placing increased emphasis on results-based management (RBM) and explicit policies and standards to measure performance. As a result, evaluation policies and strategies are being implemented in organizations around the world in an effort to increase transparency, maximize accountability and improve learning; thus, evaluation is increasingly being used as an evidence-gathering mechanism that demonstrates efficiency and effectiveness in concrete ways.

For example, within the international development context, more than 100 developed and developing countries have agreed to change the way they do business by committing to the Paris and Accra High Level Forums on Aid Effectiveness convened by the Organisation for Economic Co-operation and Development (OECD) in 2005 and 2008. The Paris Declaration is formulated around five principles: ownership, alignment, harmonization, managing for results, and mutual accountability. These principles can have a significant impact on organizational capacity, especially the principles of ownership and alignment; the principle of ownership focuses on the idea that developing countries should exercise leadership in developing and implementing their own results-oriented strategies to improve the capacity of their institutions, and the principle of alignment encourages the development and utilization of local systems and capacity to support development projects. These principles have increased the pressure on donor countries to charge recipient countries with the task of planning, implementing, and evaluating their own programs and projects

(Dabelstein & Patton, 2013). As a result, the need for local organizational evaluation capacity is growing across the developing world and, ultimately, is impacting the evaluation community and the evaluation field.

It is in this context that this thesis explores the relationship between evaluation policy and evaluation capacity building. The role of evaluation has expanded tremendously over the last few decades; evaluation has emerged as a common instrument for identifying effectiveness and efficiency in various contexts in countries and organizations that might not have used evaluation before, or that do not have the capacity to implement evaluation activities. Evaluation policies are guiding evaluation practice in these contexts, helping to identify goals, roles, responsibilities, requirements, standards and procedures, as well as expectations. Therefore, it is worthwhile to gain an understanding of the influence of such policies on evaluation capacity and the extent to which this influence is affected by different contextual factors, either as facilitators or as barriers.

To explore this context, I designed this research to provide a panoramic perspective on the relationship between evaluation policy and organizational capacity for evaluation around the world. By analyzing the evaluation policies of international organizations, involving evaluation scholars and practitioners from the United States, Canada and Europe, and engaging evaluation community members from Jordan and Turkey in discussions about the influence of evaluation policy on their capacity-building efforts, the context for the research is international. This approach helps to create connections between different contexts and provides insight into what goals policymakers want to achieve from evaluation and what outcomes local evaluators need from evaluation to help them achieve their goals, as well as into what role evaluation policy can play in connecting the two.

Overview of Methodology and Research Plan

Given that the relationship of evaluation policy to organizational capacity for evaluation is still a fairly new area of investigation that is characterized by a dearth of knowledge about how evaluation policy is implemented and how it affects practice, the design of this research is exploratory and qualitative. As will be shown in Chapter 2, the ECB knowledge base is considerably more developed than that concerning evaluation policy. I located several empirical ECB studies, some qualitative, some quantitative, and some mixed

methods. While research-based knowledge in this domain of inquiry is maturing, the same can not be said for that associated with evaluation policy. Further, the integration of these two domains of study is novel and relatively uncharted. Hence, I elected to create and implement an exploratory, qualitative design.

In order to enhance the comprehensiveness of my research, I developed a cascading design consisting of three interconnected, sequenced studies with the results of each study adding to the knowledge base and allowing understanding to develop iteratively as the overall inquiry moved forward (Lincoln & Guba, 1985). Each of my preliminary studies informed and helped to sharpen the focus for the next study in sequence and to vary the contextual circumstances for the research.

My first study (Research Question 1) is *a comprehensive and critical review and analysis of an extensive sample of evaluation policies* of various multilateral and bilateral aid agencies and national government organizations that have recently developed and/or revised their evaluation policies. This study helped focus and inform my second study (Research Questions 1, 2, & 3) comprising *one-on-one interviews* with evaluation scholars and practitioners in the evaluation field whose work has had some exposure to evaluation policy and/or evaluation capacity building initiatives. Through this study, I intended to develop deeper first-hand interpretations of the role of evaluation policy and thereby foster a better understanding of the multiple and varied dimensions that influence the role of evaluation policy. This study helped inform the third study (Research Question 3) of my research: *two focus groups with evaluation community members in two countries* in the Middle East and North Africa region (MENA), namely Jordan and Turkey. This study was designed to discuss the broad themes, issues and challenges that I identified in my interviews with scholars and practitioners in Study 2, and to gain a practical understanding of the extent to which evaluation policy has or would play an important role in the MENA region with regard to organizational evaluation capacity development.

This cascading design is comprehensive and exploratory. Despite the exploratory nature of the research, the design has a comparative element to it. Data collection for Study 1 was truly global with a wide range of developed and developing contexts represented. The sample selected for Study 2 reflects Western cultures of North America, Europe and Australasia and the results arising from this study may be situated and interpreted in this

international context. Study 3, on the other hand, tests out emergent findings in significantly distinct Middle Eastern cultural contexts. The comparison is intended to shed light on the extent to which principal findings hold in contexts where social, economic, historical, and religious norms vary quite dramatically from western countries. Such examination will help to develop our understanding of the power and role of cultural context in shaping understanding, use, and application of evaluation policy and effects on ECB.

Overview of the Thesis

The chapters of this thesis reflect the three-study approach I adopted to explore my research questions. Chapter 2, the literature review, focuses on a number of elements related to the organizational capacity to do and use evaluation and evaluation policy. Within this chapter, I develop the conceptual framework I used to guide the research. Chapter 3, “Methodology,” provides a description of the qualitative approach used herein, as well as a description of the three inter-related empirical studies in terms of data collection, data sources, and analysis involved in my research. Chapter 4, “Study 1: Evaluation Policy Analysis,” introduces the findings of a comprehensive and critical review and analysis of an extensive sample of evaluation policies of various types of organizations. Chapter 5, “Study 2,” presents the findings of the interview process with evaluation scholars and practitioners undertaken in Study 2 along with a refined framework developed based on the interview findings. Chapter 6, “Study 3,” presents findings from the two focus groups and a final version of the framework, based on the validation exercise completed in the third study. Chapter 7, “Discussion and Conclusion,” begins with a summary of my key research findings and a discussion of the influence of evaluation policy on organizational capacity. It also identifies the contributions to research and evaluation as a result of my study and concludes with reflections for further research and implications for practice.

Chapter 2: Review of the Scholarship on Evaluation Policy and Capacity Building

The purpose of this review is to situate the current body of knowledge at the intersection of ECB and evaluation policy. The review highlights some of the key theories, frameworks and empirical findings on organizational evaluation capacity and provides a summary of the current knowledge base on evaluation policy in order to set the stage for the research. In addition to surveying the landscape and consolidating existing knowledge to develop a sense of what is known and what remains to be discovered, I sought to identify gaps in our substantive knowledge about the potential impact of evaluation policy on ECB and evaluation practice.

What is Evaluation?

There are many different conceptions about the term “evaluation.” Evaluation in its more general meaning of determining the value of something has been in existence for many centuries. Evaluation as a profession based in social science disciplines, however, is relatively new. In the 1960s, the evaluation profession first emerged in the United States as distinct from its parent disciplines of education, psychology, philosophy, and sociology (Yarbrough, Shulha, & Caruthers, 2004). For this reason, evaluation is considered a “transdiscipline,” providing tools to other disciplines by drawing on a variety of social science methods (qualitative, quantitative, mixed methods) to pursue different types of information in a range of different contexts and from an array of different perspectives, but also enjoying a stand-alone status (Scriven, 1998).

Therefore, there is a rich, and sometimes competing, diversity of methods, roles, and theories of evaluation practice (Shadish, 1998). Such diversity has contributed to difficulty in defining evaluation as a unified profession, which is reflected in the many different definitions of evaluation within the field (Stevahn et al., 2005). A review of evaluation definitions by Geva-May and Pal (1999) identifies several common features of evaluation including systematic methods (Nagel & Freeman, 1975; Rossi, Lipsey, & Freeman, 1989), judgment of the merit or worth of an object (Eisner, 1979; House, 1980; Scriven, 1966), and a comparative aspect, as in assessments of two or more different approaches to a problem (Alkin & Ellett, 1990). Other common features include an aspect of feedback for

improvement, including both formative functions (Scriven, 1966) and monitoring functions (Chelimsky, 1985) and an implied role in decision making (Stufflebeam et al., 1971).

Therefore, definitions of evaluation vary in the breadth and narrowness of purpose they assign to evaluation. Some argue for an inclusive view, incorporating monitoring and performance measurement under a larger umbrella of “evaluation” activities (see, for example, *New Directions for Evaluation*, 71 [1996], entire issue). For this study, similar to others (e.g., Cousins, 2003; Weaver & Cousins, 2004) I define evaluation as systemic inquiry leading to judgment about [program] merit, worth and significance, and providing a basis for program decision making. This definition emphasizes the judgmental nature of evaluation that sets it apart from other forms of research. It also underlines that judgment in evaluation is based on systematically gathered data against a basis for comparison, which is distinct from mere monitoring activities which are predominantly descriptive. The definition also implies that evaluation results should and do inform changes in and decisions about their objects.

Evaluation Capacity Building

In recent decades, researchers have developed a number of conceptualizations and theorizations of the practices that constitute ECB. Researchers have also provided definitions of ECB that range in scope, depth, and sometimes terminology. For example, Williams (2001) and McDonald, Roger, and Kefford (2003) make the very useful distinction between evaluation capacity and evaluation capability—the former being what one holds and the latter being one’s ability to use it. McDonald et al. (2003) argue that evaluation capacity should “provide enduring organizational benefits, including a sustaining resource for producing evaluation as well as a system for encouraging and using evaluation” (p. 10). As such, “ECB is about developing the capacity to do evaluation, as well as the capacity to use it” (Levin-Rozalis, Rosenstein, & Cousins, 2009, p. 194). ECB is 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)

In other words, the development of an organization’s evaluation capacity enables the integration of evaluation practices into the organizational culture, increases evaluation use for

decision making, and continues organizational learning. This definition of ECB also recognizes the critical role of the organizational context in which the ECB process takes place, which includes structures, culture, values, material and economic resources, interpersonal and interactive components, and political dimensions such as policies, power, and influence (Mathison, 2005). This means that ECB is a process that involves changes at all levels of the organization in order to integrate evaluation practices and uses (Bourgeois & Cousins, 2008). Cousins et al. (2004) further explain that “the integration of evaluation into the culture of organizations . . . has as much to do with the consequences of evaluation as it does the development of skills and knowledge of evaluation logic and methods” (p. 101).

The review of the literature reveals that there exist a number of conceptual frameworks that emphasize different aspects of ECB. For example, Taylor-Powell and Boyd (2008) introduce a three-component framework that focuses on structuring the various strategies used in ECB in the context of extension education organizations. The authors identify professional development, resources and supports, and organizational environment as the three main dimensions of the practice of ECB within complex organizations. Along the same lines, Preskill and Boyle (2008) develop an ECB model that outlines a range of factors such as evaluation knowledge, skills, attitudes, and sustainable evaluation practice that may influence the initiation, design, implementation and impact of the ECB processes on evaluation practice. Furthermore, in a recent study, Labin, Duffy, Meyers, Wandersman, and Lesesne (2012) introduce the “Integrative ECB Model” based on a systematic review of the ECB literature. This conceptual model reflects the state of the literature on ECB by focusing on aspects of the ECB process such as the need for ECB, the activities involved in ECB and its evaluation, and the outcomes of ECB at the individual, organizational, and program levels.

Despite the increasing number of conceptual frameworks in the ECB domain of inquiry, there is a shortage of empirical research focusing on capacity-building issues *within* organizations. Consequently, there have been recent calls for the development of empirically based models that will enable a better understanding of ECB and its various components within the organization. For example, Cousins et al. (2008), Nacarrella et al. (2007), and Nielsen et al. (2011) indicate that there has been considerable focus on the role of ECB, but not much attention directed to evaluation capacity itself. Further, Cousins et al. (2014)

recently noted that “evaluation use has been underexplored and underemphasized in theory, research, and practice concerning organizational evaluation capacity” (p. 10).

In response to this shortage of empirically grounded studies in the ECB area, Bourgeois and Cousins (2008, 2013) present one of the first attempts to develop and validate a profile-based conceptual framework of the various dimensions of organizational evaluation capacity. Although the evaluation literature is replete with studies of ECB activities within various types of organizations, it was not until the development of this empirically validated framework that we had a clear description of the dimensions and sub-dimensions of organizational evaluation capacity and how it is enacted in practice. The authors show that both the capacity to do evaluation and the capacity to use it are essential in a broader evaluation capacity framework. Further, they make a very useful distinction between ECB and organizational evaluation capacity. While organizational capacity for evaluation is manifested as visible, enacted evaluation practices and processes, ECB is the process by which an organization develops its understanding and ability to undertake these practices and processes. Given this, I elected to use Bourgeois and Cousins’ (2008) framework as the underlying structure for this review.

Profile of Growth in Organizational Evaluation Capacity

Bourgeois and Cousins’ (2008) comprehensive framework is based on empirical research conducted on Canadian federal government organizations. The framework focuses on describing an organization’s capacity to do and use evaluation through six dimensions (see Appendix A). The framework has two levels: the capacity to do evaluation, and the capacity to use evaluation. At the top of the framework is the capacity to do evaluation in which the synergistic relationship between human resources, organizational resources, evaluation planning, and activities sets the stage for the ECB process within an organization. The capacity to use evaluations includes evaluation literacy, organization decision making, and learning benefits. Each of these dimensions is broken down into a number of sub-dimensions, with evaluation capacity being assessed using four levels: low, developing, intermediate, and exemplary. The framework suggests that the willingness of an organization to engage in the ECB process is the key to recognizing the learning benefits of evaluation because ECB culminates in an evaluative culture that supports continuous learning.

Capacity to Do Evaluation

Milstein and Cotton (2000) define evaluation capacity as “the ability to conduct an effective evaluation (i.e., one that meets accepted standards of the discipline)” (p. 1). This definition focuses mainly on the capacity to do evaluation. Bourgeois and Cousins’ (2008, 2013) framework identifies three dimensions of organizational capacity to do evaluation: human resources, organizational resources, and evaluation planning and activities. The human resources dimension focuses on the compensation of the evaluation unit within the organization in terms of acquiring, training, and retaining evaluators of sufficient technical and interpersonal competence to do and manage evaluation projects. It also includes the role and effectiveness of the evaluation unit’s leadership in managing the evaluation team and projects (Bourgeois & Cousins, 2013). Building an individual and organizational capacity to do evaluations is always dependent on two processes: (1) developing evaluator competencies for professional evaluation practices; and (2) planning and delivering evaluation training activities (Arnold, 2006; Garcia-Iriarte et al., 2011; Naccarella et al., 2007). LaVelle and Donaldson (2010) explain that “evaluators are made, not born, and an extended period of training is necessary to master the evaluation-specific skills and knowledge necessary to provide quality service to clients, and be socialized into the professional frameworks, standards, and ethical guidelines” (p. 10).

The second dimension of Bourgeois and Cousins’ (2008, 2013) framework focuses on organizational resources, including the available budget, ongoing data collection, and organizational infrastructure. Organizations need to allocate sufficient and stable financial resources to evaluation activities and to have a performance measurement system in place to produce the data required for evaluation studies. Organizational infrastructure, the presence of organizational evaluation policies, and strategic planning provide the essential support needed for successful evaluation. The third dimension of organizational capacity is evaluation planning and activities. Bourgeois and Cousins (2008, 2013) show that consulting with stakeholders to develop an evaluation plan, identifying evaluation needs and priorities using risk assessment processes, and gathering and disseminating critical information on an ongoing basis are crucial activities to be undertaken by evaluators.

Based on the three dimensions set out above, the capacity to do evaluations refers to human and fiscal resources needed for conducting an evaluation internally and for managing

external evaluation activities. These resources include the knowledge and skills needed for evaluation planning; training programs; instrument development; data collection and analysis; reporting and follow-up; and standards of professional evaluation practice (Cousins et al., 2008).

Capacity to Use Evaluation

The capacity to use evaluations refers to the nature of evaluation use within the organization and the extent to which evaluation use actually occurs (Cousins et al., 2004; Shulha & Cousins, 1997). The concept of evaluation use itself is multi-dimensional and refers to instrumental (e.g., decision support), conceptual (e.g., new insights), and symbolic (e.g., justification) uses. The literature reveals that the relevance, quality, and credibility of evaluations are the most important factors promoting evaluation use (Shulha & Cousins, 1997). Other factors include communication quality, findings consistent with audience expectations, timeliness in disseminating of evaluation results to decision makers, political context, and commitment to evaluation (Cousins & Leithwood, 1986). Learning from evaluation is also an important type of evaluation use that depends on the “engagement and interaction with new information and internal reflection on the meaning of that information” (Jacobson, Carter, Hockings, & Kelman, 2011, p. 55), and that, in turn, depends on sufficient organizational support.

In Bourgeois and Cousins’ (2008, 2013) framework, the capacity to use evaluations includes three dimensions: evaluation literacy, organizational decision making, and learning benefits. Evaluation literacy relates to the knowledge of evaluation within the organization, including the level of involvement of the organization’s staff in evaluation and results-management orientation. The assumption is that the participation of staff members and other stakeholders throughout the evaluation process leads to increased knowledge about evaluation and eventually to the greater use thereof (Vo & Christie, 2015; Patton, 2005, 2008).

The organizational decision making dimension is the extent to which evaluation information is integrated with organizational decision-making processes. First, organizations should include consideration of evaluation in management processes and should allocate adequate resources towards undertaking evaluation initiatives. Ultimately, evaluation

findings and recommendations are used to inform budget allocation and high-level policy and decision making (Bourgeois & Cousins, 2008, 2013). The last dimension of the organizational capacity framework is about learning benefits, with a focus on the usefulness of evaluation knowledge and information within the organization. Bourgeois and Cousins (2013) outline three types of use that are related to learning benefits in the organization: instrumental use, conceptual use, and process use. Instrumental use occurs when the program or policy goes through substantial changes as a result of the evaluation and when these effects are clear and visible (Neuman et al., 2013). Conceptual use relates to broader changes in program staff attitudes despite an absence of practical change at the program level (Patton, 2001). What may occur instead is the clarification of the program logic model, the identification of measurement indicators or the production of lessons learned (Patton, 2001). Finally, process use stresses the importance of stakeholder involvement and learning at all stages of evaluation (Patton, 1998). The benefits of process use are presumed to be long-lasting, outweighing those of a recommendation report, which can quickly lose its timeliness. What sets this type of use apart is its emphasis on “learning how to learn” (Patton, 1998, p. 226). In addition to the instrumental, conceptual, and process use identified in the framework, considerable theoretical and conceptual work on use has led to a variety of distinctions between other types of use, such as between symbolic use and process use and misuse (e.g., Cousins & Leithwood, 1986; Shulha & Cousins, 1997; Weiss, 1977, 1979).

The empirically grounded framework describes in detail the dimensions of the two channels of organizational evaluation capacity and how they develop over time. As such, it enhances our understanding of the potential impacts of organizational improvement initiatives by showing the aspects of the two types of organizational capacity. The framework also serves as a valuable tool to generate serious discussion and debate about evaluation capacity and optimal strategies and policies for improving it. Although it is not part of Bourgeois and Cousins’ (2008) framework, evaluation policy is inextricably linked with the integration of evaluation practice into the organizational culture for organizational development and learning.

Understanding Policy

The literature on evaluation contains many definitions of evaluation and evaluation capacity, but almost none of evaluation policy. Regrettably, very few of the works found in the literature review for the present study included a definition of the term. Since one of the main purposes of this study is to empirically explore the role of evaluation policy in relation to evaluation practice, it will be useful to start from a very broad and inclusive definition that leaves room to move in many different directions.

The common dictionary definition of “policy” is (a) “a definite course or method of action selected from among alternatives and in light of given conditions to guide and determine present and future decision or (b) a high-level overall plan embracing the general goals and acceptable procedures” (Merriam Webster). This definition means that policy involves an intentional and deliberate process that considers both needs and alternatives in order to arrive at a decision. It also involves the exercise of power to determine action, rather than being the mere expression of an idea. The action is goal-directed and is to be repeated in the future, rather than taking place only in a single instance. The standard dictionary definition, however, does not specify how explicit a plan must be to be considered a “policy.” Consequently, it is possible to consider examples of policy at different levels of formality and informality, rigidity and flexibility, ranging from policies that are highly formal, explicit, and rigid to policies that are informal, implicit, and changing.

There is an enormous amount of writing about policy across academic disciplines in the social sciences, though exactly what policy means is largely assumed rather than explained. Bauer (1968) provides a very useful explanation:

Various labels are applied to decisions and actions we take, depending in general on the breadth of their implications. If they are trivial and repetitive and demand little cogitation, they may be called routine actions. If they are more complex, have wider ramifications, and demand more thought, we may refer to them as tactical decisions. For those which have the widest ramifications and the longest time perspective, and which generally require the most information and contemplation, we tend to reserve the word policy. (pp. 1–2)

This useful distinction between routine actions, tactical decisions, and policy highlights the cognitive and practical hierarchy among everyday procedures and a deeper, long-term vision. In other words, policy creates the frames of meaning within which everyday practices happen, or what Bauer (1968) describes as “parameter-shaping acts.”

The assumption that underlies the mainstream definition of policy is that it is a purposive course of action (Anderson, 1997). For example, Mayer and Thompson (1982) explain that policy is a declaration that defines the intention of an organization's goals and priorities, outlines roles, rules and procedures, and creates a framework within which the administration and staff can perform their assigned duties. Similarly, Titmuss (1974) defines policy as "the principles that govern action directed towards given ends. The concept denotes action about means as well as ends and it, therefore, implies change" (p. 23). It is therefore a course of action designed to achieve particular results (Bridgman & Davis, 2004). Colebatch (2009) provides a different perspective and investigates policy as a practical concept. The author notes that policy suggests expertise because it concerns a particular problem area, implying knowledge of the problem and what might be done about it. When new policy concerns appear, they are driven initially by widely shared principles that are developed into a body of specialized policy over time. This has given rise to what Colebatch (2009) calls "adjectival policy": the development of expertise focused on a specific area of practice to be governed or guided by a policy, such as evaluation policy.

Although there is no consensus regarding the definition of policy, what these definitions and theorizations share is the recognition that policy influences action and that it is a matter of means and ends and the relationship between them. For this study, I define policy as the standard ways of doing things that influence decisions, directions, and actions to reach a specific goal or goals.

It should be noted that policies do not and cannot be formulated and implemented in a vacuum. Policies are influenced by the contexts in which they are developed, such as the organizational context and information needs (content), the nature of the policy process, and the actors involved in the formulation and implementation process. The section below discusses the dynamic components of policy development and implementation.

Policy Formulation and Implementation

Several theories exist on the key components of policy formulation and implementation and on ways in which to judge the success of such processes. While policy formulation involves developing strategies for dealing with issues that have been placed on the political agenda, policy implementation refers to the mechanisms, resources and

relationships that link these strategies to actions in order to achieve a specific goal (Bhuyan, Jorgensen, & Sharma, 2010). More specifically, policy implementation means to carry out, accomplish, fulfil, or complete a given task (Bardach, 2008).

The development of policy involves several stages from inception to completion, as are described in various policy stage frameworks. These stages are agenda setting, formulation, legitimation, implementation, and evaluation. The policy stages framework is referred to in a number of ways and has been variously called the “linear model,” the “heuristic stages model” or the “policy cycle model” (see, for example, Anderson, 2011; Brewer & DeLeon, 1983; DeLeon, 1999; Knoepfel et al., 2011; Smith & Larimer, 2009).

The *agenda setting stage* is the first critical step in the policymaking process that is concerned with deciding what is to be decided. It starts with the emergence and acknowledgment of a problem that triggers a collective need or dissatisfaction for which a solution is needed. Such a problem exists when there is a difference between the current and the desired situation (Knoepfel et al., 2011). For a policy to be placed on the agenda of a given organization, the key actors must acknowledge that a situation is problematic, identify and clarify the problematic aspects of the situation, propose solutions, and engage in activities and discussions to identify alternatives to address the problem (McCool, 1995).

The *policy formulation stage* presupposes the identification of the cause of the problem and is concerned with examining the various options that are considered to be possible solutions. This is a critical stage of the policymaking process because it involves the selection of objectives, instruments, and procedures to be implemented to address the issue under consideration (Smith & Larimer, 2009). It should be noted that it is during this stage that different actors strive to gain priority for one specific interpretation of both the problem and its solution and that power relationships crystallize, thus determining the direction that the policy will take.

The *legitimation stage* is the stage of policymaking during which decisions that favour one or more approaches to address a given problem are made. The final selection made from the various proposed policy options depends on two major factors. The first factor is the feasibility of the policy, and this largely depends on the availability of the required resources to effectively implement the policy, including personnel and material resources (Sabatier & Mazmanian, 1995; Swinburn et al., 2005). The second factor is the capacity of

the policymakers to set strategic directions, to weigh and assess the implications of the various policy alternatives, and to make appropriate use of knowledge in policymaking (Bakvis, 2000; Howlett & Lindquist, 2004; Parsons, 2004).

The *implementation stage* refers to the stage between a decision and operations. It is the challenging next step after the decision, and it involves efforts to put in place what has been decided with regard to the policy. Several factors determine the effects of a policy and how well it is implemented to achieve its objectives. These factors include the type and complexity of the problem addressed, the magnitude of the expected change, the groups targeted by the policy, the human and financial resources devoted to the implementation, and the administrative structures that are available to support the implementation of the policy (Sabatier & Mazmanian, 1995). It should be noted that the decision on a specific course of action does not guarantee that the actions on the ground will strictly follow the policymakers' goals and objectives. "To decide" does not necessarily mean "to do." For organizations, a decision requiring implementation demands both commitment and capacity for execution. In the case that a policy is externally imposed on an organization, a third demand arises—that is, communications. It is argued that policy implementation most straightforwardly concerns how to bring together commitment, capacity, and communications so as to convert the decisions into actions (Jann & Wegrich, 2007).

The *evaluation stage* refers to assessing the impact of the policy in delivering the desired result. During this stage, a policy is evaluated to verify the extent to which its implementation and effects are aligned with the objectives that were defined during the formulation stage (Howlett & Ramesh, 2015). In this context, policymakers make a decision as to whether the policy should be revised. It is critical to keep in mind that evaluation is not restricted to a particular stage in the policy cycle; instead, it is an embedded part of the entire policymaking process and debate (Jann & Wegrich, 2007).

This conceptualization of policy stages is useful for grasping the process by disaggregating it into more specific elements—though, as Daniell, Marton, and Insua (2015) recognize:

The policy cycle—and rational decision-making processes more generally—are often criticized as rarely reflecting what happens in reality. . . . However, the stages noted above still prove useful in practice to orientate needs and thinking around different forms of policy (p. 3).

The complexity of looking at policy as a process is that these stages typically do not occur sequentially. Instead, they are interrelated, sometimes iterative, and some may even occur simultaneously. Moreover, the multiple actors with varying interests and power relations can influence the process at any particular stage. In fact, power continues to shape the parameters of relationship and, ultimately, the process and outcomes of policy formulation and implementation.

Policy and Power

The role of power is central to policy. The reason why some policy issues get on the agenda and move to the formulation stage while others disappear can be explained through the understanding of the concept of power. The way in which power is manifested—the form that it takes—throughout the policy process influences the input and the outcome of the policy (Gaventa, 2006). Power is a dynamic and relational concept (Foucault, 1980) that informs the relationships of the actors who are involved in the formulation and implementation of a policy as well as the broader context within which the process takes place (Gaventa & Cornwall, 2006). Therefore, power is not something that is possessed by an individual or a higher authority but rather something that structures and circulates knowledge at different levels (Foucault, 1980). As Gaventa (2003) explains, “power is diffused rather than concentrated, embodied and enacted rather than possessed, discursive rather than purely coercive, and constitutes agents rather than being deployed by them” (p. 1).

In my review of the literature, I identified three types of power that help to explain its dynamic nature within the policy-process context and differentiate between the levels at which it operates. The three types of power are relational, dispositional, and discursive. The relational power (also referred to as “power as capacity”) frames the relationships among the various actors within the policy context and creates categories of people who can participate in the policymaking process (Veneklasen & Miller, 2002). This type of power also frames the negotiation and decision making among various actors and places the focus on the rules of inclusion in and exclusion from the process, on competing interests, and on who can or will speak for whom (Goehler, 2000). Power is always constituted and exerted in social relationships; therefore, relational power implies actors, resources, outcomes, and interactions among them. Within this type of power are transitive and intransitive power

(Goehler, 2000). While transitive power refers to the power struggles that occur when some actors achieve outcomes against the will of others, intransitive power exists in the joint practices of actors and encompasses the ensemble of relationships that constitute a group of people as a community. This distinction is important in the context of policymaking, as both forms of relational power are linked to policymaking. In policymaking, actors struggle for certain individual outcomes, and a victory for one may be a loss to another (transitive power). At the same time, policymaking aims to achieve a common benefit for the entire community affected by the policy (intransitive power).

The second type of power is dispositional power, which shapes the organization's capacity to act (Clegg, 1989). Through this type of power, the positions of various people in an organization determine what they may achieve in terms of relational power. This type of power is affected by two major factors within the organization: rules and resources. Organizational rules define and legitimize the respective positions that agents in an organization may occupy, and the allocated resources determine the relative dependency of an agent in a particular position (Clegg, 1989). It should be noted that dispositional power allows certain people the authority to reinterpret the meaning of the rules. Because of the power that the rules give them, they possess greater freedom to activate them according to their own interpretation (and sometimes interests) than do those people who are in positions that do not allow them such autonomy. This type of power is associated with biases and influences whose voices are heard and whose agendas are included in the policymaking process, especially at the agenda setting and formulation stages (Clegg, 1989).

The third type of power is discursive. It links the relational and the dispositional power and refers to the internalized norms and values that guide practice and influence signification, legitimization, and domination in discourses (Giddens, 1984). Discourses link the relational power to the production of knowledge (Foucault, 1980), with influence on all stages of the policymaking process. Foucault's interpretation of the term "discourse" is broad and not restricted to texts or sets of thoughts. Instead, discourse refers to practices that exhibit consistency. Discourse "is not a consciousness that embodies its project in the external form of language; it is not a language plus a subject to speak it. It is a practice that has its own forms of sequence and succession" (Foucault, 1972, p. 169).

Thus, discourses can assume textual forms, but texts are only a part of what constitutes discourse. Certainly, the process of producing texts (such as policy documents) takes place within social practices and political governance structures that influence what is included and what is excluded in the process of forming discourse (McNay, 1992). For Foucault (1980), discourse not only structures knowledge but also creates and legitimates it: “In any society, there are manifold relations of power which permeate, characterize and constitute the social body, and these relations of power cannot themselves be established, consolidated nor implemented without the production, accumulation, circulation and functioning of discourse” (Foucault, 1980, p. 93).

At any one time during the policymaking process, there will be a multiplicity of discourses, some competing or in tension with one another, and others in relationships that are broadly reinforcing. It is this multiplicity that opens up spaces for resistance:

We must make allowances for the complex and unstable processes whereby discourse can be both an instrument and an effect of power, but also a hindrance, a stumbling block, a point of resistance and a starting point for an opposing strategy. Discourse transmits and produces power; it reinforces it but also undermines it and exposes it, renders it fragile and makes it possible to thwart it. (Foucault, 1978, p. 101)

Thus, discourses are embedded in social systems and, because of their capacity to be transformed, discourse should not be conceptualized as a fixed entity but rather as a window to a partial and situated “reality.” Therefore, various forms of discourse enable us to make sense of a given social reality and to understand how power and knowledge are interlaced and articulated through social constrictions. The basic premise of discursive power is that the ways in which people think and talk about a subject influence and reflect the ways in which they act in relation to that subject (Foucault, 1980). Therefore, the analysis of discourse (e.g., policy) provides an opportunity to explore power relations and to understand the production and consumption of knowledge. Although less obvious and invisible, discursive power shapes our values and beliefs and ultimately influences how we think and what we think about.

When it comes to policy, problem definition shapes what issues get on the agenda, and what specific course of action is or is not taken. How policymakers interpret current and past events shapes the ways in which they define a problem and helps to frame and label issues for decisions (Rochefort & Cobb, 1994). Labelling an issue informs the kind of

attention that the issue attracts and sets the stage for decision making (Peters, 2005). Therefore, what is usually more important in influencing the policy agenda setting and the formulation stage is the interpretation of issues (Princen, 2011) and the subsequent labelling of such issues. Different policy actors present different explanations of the nature of a particular problem (Portz, 1996) and use different negotiation tools, such as resources or information, to make a case and persuade others.

Espoused Policy and Policy in Use

Concern with the relationship between espoused policy and policy in use is, of course, not new as the notion of espoused theory (how organizations think they operate) and theory in use (how organizations actually operate) was introduced by Argyris and Schon (1974) in their book on organizational learning. The extent to which organizations' policies are closely aligned with their actual practice is a question of implementation. The review of the literature reveals that research on policy implementation has been a dominant discourse among policy scholars since the 1970s when Pressman and Wildavsky (1973) brought the issue of policy implementation to the forefront. Their influential study offers significant insights into the practical implications of the failure of a development program. The authors show that, after years of considerable expenditure and significant effort on the part of many stakeholders who apparently agreed on a program's importance and usefulness, it became clear that the program had resulted in almost no useful outcomes. In reflection on this project, Pressman and Wildavsky (1984) define implementation in terms of fulfilling, carrying out, and accomplishing a task as it is laid out in official policy documents. Similarly, O'Toole (2003) refers to policy implementation as the connection between the expression of organizational intention and the actual result.

The review of the literature on policy implementation also shows that it is difficult to identify specific factors or conditions that facilitate successful implementation since so much depends on the context within which a policy is implemented (i.e., the social, political, and economic context) (Payne, 2008). Meier (1999), a policy scholar, comments sarcastically on the enormous number of variables affecting implementation that one can find in the literature and proposes that "any new policy implementation scholar who adds a new variable or a new interaction should be required to eliminate two existing variables" (p. 6). Nonetheless, some

scholars have identified important factors that can be leveraged to minimize the gap between policy and practice and thus increase effectiveness. For example, Elmore (1978) identified four major factors for effective implementation: (1) clearly specified tasks and objectives that accurately reflect the intent of a policy; (2) a management plan that assigns tasks and performance standards to subunits; (3) an objective means of measuring subunit performance; and (4) a system of management controls and social sanctions sufficient to hold subordinates accountable for their performance (p. 195). On the other hand, Kaufman (1973) cites three reasons for policy implementation failures: (1) the inability of implementers to understand policy objectives; (2) the lack of capability or resources to follow policy directives; and (3) disagreement with policy directives. This emphasis on individuals' commitment and organizational conditions implies that implementation is a process that is closely related to capacity within the organization.

What is Evaluation Policy?

Trochim et al. (2009) identify evaluation policy as being “of considerable importance, especially in relation to the limited amount of attention it receives as a general topic in the mainstream evaluation literature” (p. 3). In fact, the results of a literature search show that it was considered a topic of interest in 2008 and 2009, with eight publications appearing in the journal *New Directions for Evaluation* in 2009 but none thereafter. Trochim, Mark, and Cooksy's (2009) work has started to define the limits of the domain and to provide some direction for understanding evaluation policy and its dimensions. In “Evaluation Policy and Evaluation Practice,” Trochim (2009) defines evaluation policy as “any rule or principle that a group or organization uses to guide its decision and actions when doing evaluation” (p. 16). This definition allows room for different levels of application, from the highest level of general principles to the lowest level of specific directions. It also allows room for policies that are not directly concerned with evaluation but influence evaluation-related actions and decisions. Based on this definition, evaluation policy is inherently collective and could be written and explicit or unwritten and implicit. The definition expands the scope of the construct to include rules that influence evaluation decisions and action but are not identified through an intentional process.

Why Does Evaluation Policy Matter?

As evaluation contexts become ever more complex, evaluation practice evolves to accommodate different roles, purposes, and requirements for different organizations. Within this context, evaluation policy governs the way evaluation is conducted and the use of evaluation results. It carries an organization's intentions to be accountable to stakeholders, to support organizational learning, and to use evidence for decision making.

Evaluation policy has the potential to influence decision making in a range of ways. Policies about when to evaluate outcomes, whether early or late in a program's development, can help determine whether the program is able to demonstrate success. Policies about what constitutes success can determine whether a program or policy is considered to be successful. Policies about how to evaluate—which methods and approaches to use—can make the difference between a finding of impact or no impact. Policies about whether external professionals or internal program administrators will conduct evaluations, and where internal evaluators are situated in the organization, can affect whether evaluation results are used. Policies that link evaluation results to funding can increase accountability and better use of resources. Policies about how much influence to give evaluation results can determine which programs policymakers choose to continue, improve, or terminate (Chelimsky, 2009; Datta, 2009; Mark, 2009; Trochim, 2009).

Policy thus sets an important frame of reference for how funders and evaluators define, implement, report, and use evaluations. These policies exist within public, private, and nonprofit sectors. The GPRA and the Treasury Board policy on evaluation, for example, are both evaluation policies. It should be noted that evaluation policy is created and applied in a specific political, economic, and cultural contexts, whether at the organizational, national, or global levels. For instance, in organizations where the political context threatens the credibility or accuracy of evaluations, the establishment of an evaluation policy would provide the needed protection for the evaluation function. Stern (2009) shows in an article about evaluation policy in the European Union (EU) that the EU's political context, which includes 27 member states, creates a need for evaluation to meet international transparency and accountability requirements; thus, several policies related to evaluation use and its links to decision making have been created and implemented. Further, Leeuw (2009) describes how evaluation policy plays an important role in an economic context. He describes the

“reconsideration studies” conducted in the Netherlands in the 1980s by internal ministry staff who required that all evaluation reports have a section describing the potential consequences for goal achievement if there was a 20 percent cut in the budget allocated to an intervention. These examples show how evaluation policy is influenced by and can influence the different contexts in which it is implemented.

Evaluation Policy as Part of the Evaluation System

In the greater scheme of things, evaluation policy is a significant factor that can help determine evaluation practices of an organization within a broader *evaluation system*. According to the Organisation for Economic Cooperation, Development Assistance Committee (OECD DAC, 2004), evaluation system refers to “the procedural, institutional and policy arrangements shaping the evaluation function and its relationship to its internal and external environment . . . [which] includes the evaluation function’s independence, the resources it relies upon and, not least, cultural attitudes to evaluation (p. 2).

The evaluation system is of key importance within the organization because it influences the demand for evaluation as well as its use, including the dissemination and integration of evaluation results. Dahler-Larsen (2006) argues that an evaluation system acts as “a continuous stream of information that plays an integrated role in the functioning of an organization to produce evidence-based data” (p. 1). To a great extent, this depends on the policy of evaluation and the existence of supportive organizational design and staffing structure that allow the flow of information. Leeuw and Furubo (2008) indicate that an evaluation system must represent an organization’s internal consensus about the meaning of evaluation, the purpose it serves, and the type of knowledge it produces. The authors also argue that the responsibility for evaluation must rest with the organization, rather than being outsourced. Such practices ensure permanence to the evaluation activities within the organization and support its capacity for evaluation.

Seegerholm (2003) further observes that internal organizational aspects including how the organization is structured, power relations, and staff’s views on and knowledge of evaluation influence evaluation practices within the organization and the way it interacts with its context. On the other hand, Chelimsky (2009) noted the influence of culture and the surrounding context of the evaluation system on evaluation practice. She argues that the

culture, bureaucratic system, and governmental structures can constrain evaluation practices of an organization and that more intentional design of evaluation policies is required to address systematically the issues generated by such contextual factors.

In addition to conceptualizing evaluation policy as a critical factor within the broader spectrum of organizational system, structures, and context, it is also important to view it as a structure in and of itself within the organization. The following section describes the categories of evaluation policy to further our understanding of its key components.

Categories of Evaluation Policy

Trochim (2009) introduces a thoughtful methodological framework for evaluation policy that includes a taxonomy of evaluation policies and a structure that identifies the interrelationships between policies and practices. The framework includes eight categories of evaluation policy and the evaluation issues and questions they address. These categories are as follows: evaluation goal (e.g., evaluation definition); evaluation participation (e.g., who can be involved in evaluation); evaluation capacity building (e.g., human resources); evaluation management (e.g., requirement for evaluation); evaluation process and method (e.g., evaluation approaches recommended or required); evaluation role (e.g., responsibilities for evaluation); evaluation use (e.g., how evaluation results will be used); and meta-evaluation (e.g., meta-evaluation to assess the implementation, quality, and utility of the evaluation). Trochim's (2009) framework provides a comprehensive tool to understand the categories of evaluation policy.

Leeuw (2009) used this framework to assess the state of evaluation policy in the Dutch government. He was able to show that there exist some gaps in the governmental evaluation model in terms of evaluation methods and participation policies. Leeuw (2009) argues that such gaps can be attributed to the multiple and separate functions for evaluation in the country. Such an example indicates that Trochim's framework has great potential for identifying gaps in evaluation policy and for refining our understanding of the reasons for which the gaps exist, which would be of considerable usefulness in focusing efforts to develop organizational evaluation capacity.

Conceptual Framework

What is the role of evaluation policy in enhancing organizational evaluation capacity? Integrating Bourgeois and Cousins' (2008, 2013) empirically grounded framework for organizational evaluation capacity (Appendix A) and Trochim's (2009) theoretical framework for evaluation policy provides a useful tool to answer this question. The comprehensive dimensions and sub-dimensions of Bourgeois and Cousins' framework enable a broad understanding of organizational evaluation capacity, and the evaluation policy categories outlined in Trochim's framework allow for a broader conceptualization of the ways in which these dimensions are shaped, reinforced, or impeded by evaluation policy.

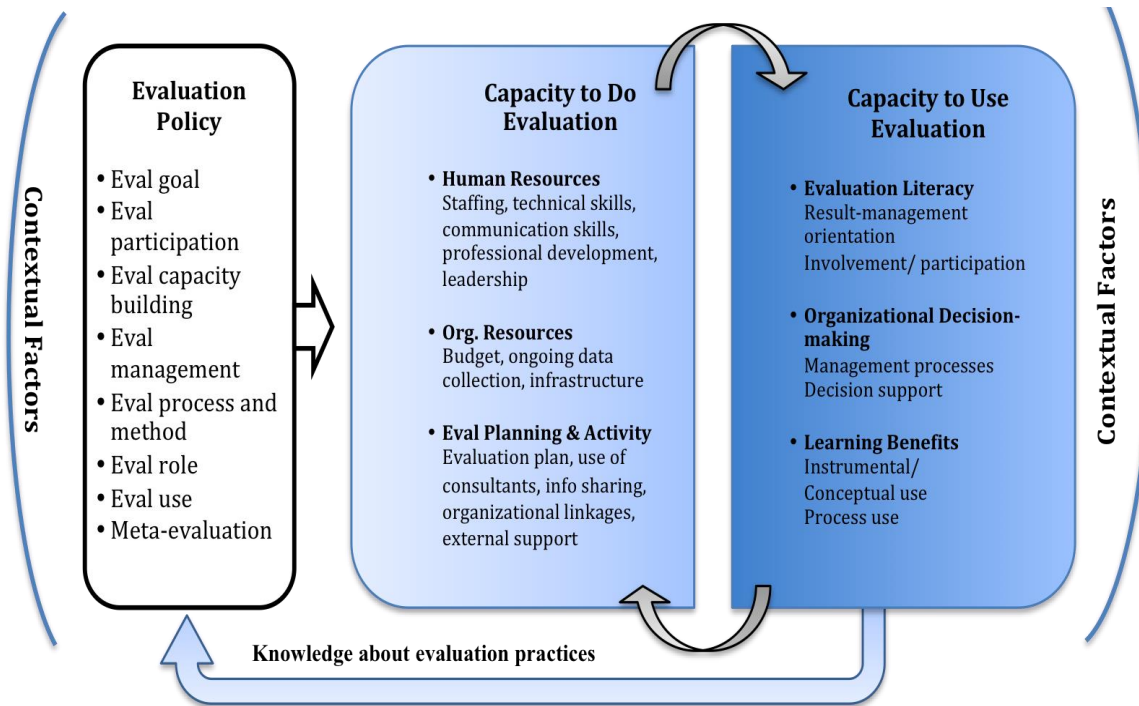


Figure 1. Conceptual framework describing the relationships between evaluation policy and organizational capacity to do and use evaluation.

By integrating the two frameworks, I developed the conceptual framework set out in Figure 1. This framework supports and informs the present study on the role of evaluation policy in organizational evaluation capacity by explicating the key constructs to be studied as well as the presumed or suggested relationships among them (Miles & Huberman, 1994). The first aspect of the framework is the eight categories of evaluation policy suggested by Trochim (2009). These evaluation policy categories are depicted as influencing and shaping

the organizational evaluation capacity process, which corresponds to the six dimensions of organizational evaluation capacity set out by Bourgeois and Cousins (2008, 2013). While the evaluation policy directly informs the dimensions and sub-dimensions of organizational evaluation capacity to *do* evaluation (human resources, organizational resources, evaluation planning and activity), it also sets the stage for the process within which an organization develops its capacity to *use* evaluation. The third panel of the framework lays out the dimensions of the capacity to use evaluation, including evaluation literacy, organizational decision making, and learning benefits. The capacity to do and the capacity to use evaluation are interactive and inform one another, as is indicated by the two arrows that connect them. The arrow from the last panel to the evaluation policy indicates that evaluation policy and evaluation use are connected in such a way that, as the organization develops its capacity to use evaluation, it accumulates knowledge about effective evaluation practices that will allow for the subsequent revision and improvement of evaluation policy. The framework also recognizes the fact that ECB is context dependent and that contextual factors are likely to influence all aspects of the process as facilitators or barriers. Context here is defined as

the combination of factors accompanying the implementation and evaluation of a project that might influence its results, including geographical location, timing, political and social climate, economic conditions, and other things going on at the same time as the project. It includes the totality of the environment in which the project takes place. (Thomas, 2004, p. 11)

My integration of the respective frameworks of Trochim (2009) and Bourgeois and Cousins (2008, 2013) provides a different and useful lens through which to broaden our understanding of organizational evaluation capacity. When considered together, the two frameworks inform one another. Moreover, by adding the contextual aspect, the integrative framework provides a holistic picture of what to consider when studying organizational evaluation capacity and shows the dynamic among the different aspects of the process. The framework draws on the research of leaders in the evaluation field and acts as a bridge connecting their efforts and expertise. It thereby contributes to the consolidation of knowledge in order to create a basis for further advancements in the evaluation field. Most importantly, the framework identifies the most significant elements to investigate when conducting an empirically based study that seeks to add to the knowledge base and fill gaps in our knowledge about the role of evaluation policy in organizational capacity for

evaluation. For these reasons, the framework presented above is integral to this research study, guiding instrument development, data collection, analysis and interpretation, and reporting. Nevertheless, the research methodology is open to refinement based on emerging observations, constructs, and relationships (Baxter & Jack, 2008; Miles & Huberman, 1994).

Chapter 3: Methodology

Overall Research Design

The research design is the logical connection between the data to be collected and the conclusions to be drawn on the one hand and the research questions on the other hand (Stake, 1995). Given that the relationship of evaluation policy to organizational capacity for evaluation is still a fairly new area for investigation and is therefore characterized by a dearth of knowledge, the research design I developed for this study, as depicted in Figure 2, consists of three interconnected studies with the results of each study adding to the knowledge base and allowing understanding to develop iteratively as the overall inquiry moves forward (Lincoln & Guba, 1985).

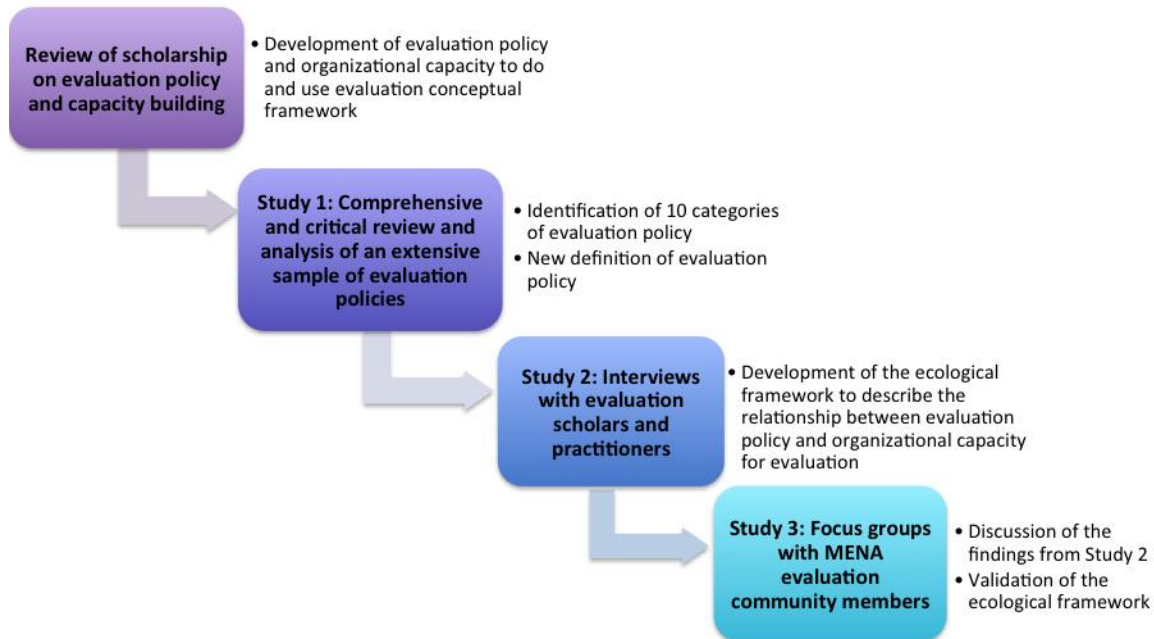


Figure 2. Summary of methodology and logical integration of studies.

The research started with a review of the scholarship on evaluation policy and evaluation capacity building to highlight some of the key theories, frameworks, and empirical findings and to provide a summary of the current knowledge base in these two areas to set the stage for the investigation.

Study 1 (Research Question 1) of this inquiry is a comprehensive and critical review and analysis of an extensive sample of evaluation policies of various multilateral and

bilateral aid agencies and national government organizations, some of which have recently developed or revised their evaluation policy. The purpose of this study is to scan the landscape and develop and present an understanding of the current trends in evaluation policy and thus to answer the first research question on the role of evaluation policy in organizational ECB. This study also informed the instrument development for the second study.

Study 2 (Research Questions 1, 2 & 3) consists of one-on-one interviews with evaluation scholars and evaluation practitioners whose published work has focused on evaluation policy and/or evaluation capacity building initiatives. The interview data apply to all three primary research questions of the thesis; the study is intended to develop deeper first-hand interpretations of the role of evaluation policy and thus to foster a better understanding of the multiple and varied dimensions at play within the organizational context that influence the role of evaluation policy. It should be noted that Study 1 and Study 2 focus largely on espoused evaluation policy although perspectives on evaluation policy in action emerged in the interview study. The findings of these studies informed the instrument development for the third study.

Study 3 (Research Question 3) consists of two focus-group discussions with evaluation community members and evaluation professionals in a particular region in which multilateral and bilateral aid and governmental organizations are active. The focal region is the Middle East and North Africa (MENA), and the study involved two countries: Turkey (transitional economy) and Jordan (developing economy). I selected this region because it is the region of my heritage and upbringing and it represents a distinct non-western context. Despite these countries' relatively limited history with evaluation, interest in developing organizational evaluation capacity in the region is high, as is reflected in the establishment of the MENA Evaluation Network (EvalMENA) in 2011; its recent participation in the International Year of Evaluation in 2015; and, more recently, the establishment of the Jordanian Development Evaluation Association (EvalJordan) in 2014 and the Turkish Monitoring and Evaluation Society (TMES) in 2013. This study is intended to answer the third research question as it is focused on gaining an understanding in a practical sense of the contextual facilitators and barriers that influence the role of evaluation policy in enhancing the integration of evaluation into organizations; this by looking at the extent to which

evaluation policy has played or is likely to play an important role in MENA with regard to organizational evaluation capacity development. The study also sought to discuss the findings from Study 2 with members of the evaluation communities in the region.

The design of this research has benefited from the Validity Network Schema (VNS) framework presented by Brinberg and McGrath (1985), a framework that captures the process of social science research (see also McGrath, 1994). The VNS shows that the social science research process involves bringing together three interrelated domains: (a) the *substantive domain*, which refers to the content that is of interest to the researcher; (b) the *conceptual domain*, which gives meaning to the content; and (c) the *methodological domain*, which refers to the techniques or procedures by means of which those ideas and contents can be studied. According to McGrath (1994), the research process, like a three-legged stool, involves sampling from these three domains to understand a given phenomenon and the relations among its elements within a context or embedding system. The framework describes three stages of research and the activities involved in each of them; these include prestudy (Stage 1), study (Stage 2) and poststudy (Stage 3) activities and alternative paths available to researchers. Within this framework, validity has three distinct meanings that are linked to each stage of the research. This theoretical conception serves as the basis for the research process employed in this investigation. This approach to logical inquiry is popular in the business field, especially in marketing, economics, and organizational studies (see, for example, Brinberg & Hirschman, 1986; Greenwood & Suddaby, 2006; Holm, 2007; Hornsby, Kuratko, & Montagno, 1999).

Figure 3 shows the research path selected for the present investigation. At Stage 1, the review of the scholarship on evaluation policy and ECB led to the development of the evaluation policy and of the organizational capacity to do and use evaluation conceptual framework illuminating key elements and the preserved relations among them. This step involved sampling from the *conceptual domain* in the Brinberg and McGrath framework (1985) and was undertaken in the absence of theoretical frameworks that explain the potential relations between evaluation policy and organization evaluation capacity.

At Stage 2, the conceptual framework developed in Stage 1 was used to guide instrument development and qualitative data collection (the *methodological domain*) in Study 1 and Study 2. Combining the conceptual and methodological domains resulted in a research

design that was then applied to understand the role of evaluation policy in organizational capacity for evaluation in practice (the *substantive* domain). At Stage 3, the poststudy involved the verification and delimitation of the set of findings that resulted from Stage 2 activities using repeated sampling of the methodological domain.

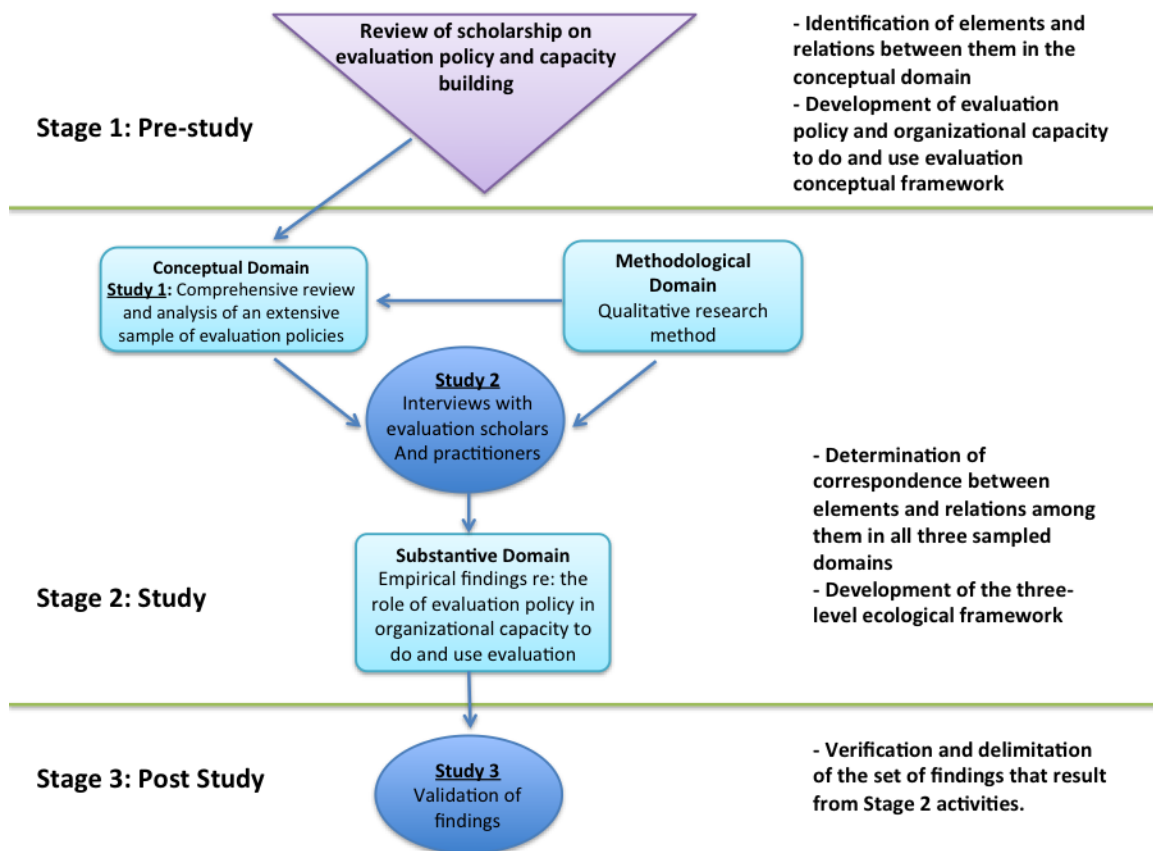


Figure 3. Conceptual basis of research plan.

This research design is consistent with the rationale of my qualitative approach and allowed the research itself to “unfold, cascade, roll and emerge” (Lincoln & Guba, 1985, p. 210). In this way, the research design evolved as a living entity that took shape throughout the research process. Therefore, the prestudy, the study and the poststudy stages of the research design are interrelated and overlapped substantially and allowed me to answer the research questions on the role of evaluation policy, its influence on organizational capacity, and contextual factors. It should be noted that, even though a structure was imposed on the data-collection procedures (i.e., the conceptual framework in the prestudy stage), the research process allowed for the emergence of unanticipated phenomena or patterns in the data (hence

the emergence of the ecological framework in Stage 2). This approach to qualitative empirical research is based on that developed by Miles and Huberman (1984), which has been used in evaluation and wide range of research studies in education and the social sciences.

To follow is a detailed description of each study and the methods used in each, but first I will comment on the epistemological orientation that I bring to the research.

Epistemological Orientation

I am a woman of Middle Eastern heritage and have a deep understanding of the historical, social, economic and political complexities of this particular region of the world. Despite solid early educational experiences in my home context, my graduate training has been exclusively located in Canada. It has been through my graduate training that I have developed my epistemological orientations that underlie systematic inquiry.

The research design for this thesis is grounded in a predominantly logical empiricist approach. The use of existing, preordinate knowledge on both evaluation policy and evaluation capacity in the integrative framework that guided this research through the process of reviewing existing evaluation policies in Study 1 and through the interviews that took place in Study 2 assumed a mostly logical empiricist stance. That is to say that I used the conceptual framework laid out in Chapter 2 to guide instrumentation, data collection, analysis, and interpretation, while at the same time remaining open to emergent concepts and constructs. In remaining open in this way, particularly in Study 2, I acknowledge and respect the importance of context and of the role of the social construction of knowledge. The methods used for the focus groups in Study 3 were even more reflective of social constructivist practices. Social constructionism, which is concerned with the importance of context and the inclusion of multiple participants in the co-construction of knowledge, is focused on uncovering how individuals and groups participate in the construction of their perceived social reality (Greene, 1994). It involves looking at how social phenomena are developed, institutionalized, made known and converted into tradition (Lincoln & Guba, 1985). So, in essence, my epistemological stance for the research can be framed as a blend of logical empiricism and social constructivism, perhaps leaning slightly more towards the former.

My research draws on shared understandings of the role played by evaluation policy in organizational evaluation capacity. This process involved the social construction of knowledge through the sharing of thoughts and experiences during the interviews and focus-group sessions while acknowledging the fact that the categories of evaluation policy can be identified in a quasi-objective manner beyond the direct experience and environments of the participants. This resulted in a series of findings that are transferable across organizations that might not have been involved in the study. Finally, while I recognize that “human meanings and understandings and intentions are worked out within the frameworks of social structures” (Miles & Huberman, 1994, p. 4), I believe that social phenomena exist not only in the mind but also in the objective world and that there are reasonable relationships to be found among them. Given the scope of this research and its focus on increasing understanding of the relationship between evaluation policy and organizational capacity using existing knowledge in building the research guidance framework, it follows that the main theoretical approach continues to be based on logical empiricism.

Study 1: Evaluation Policy Analysis (Research Question 1)

The purpose of this study is to scan the landscape and understand the current trends in evaluation policy in relation to organizational capacity to do and use evaluation. This study consists of a comprehensive and critical review and analysis of an extensive sample of evaluation policies of different multilateral and bilateral aid agencies and government organizations that have recently developed or revised their evaluation policies. This study is intended to answer the first research question (What is the role of evaluation policy in the organizational capacity to do and use evaluation? To what extent is the development of organizational capacity to do and to use evaluation espoused in existing organizational evaluation policies?). The study is important (i) for obtaining valuable information about the current trends in evaluation policy development and its connection with evaluation capacity; and (ii) for informing instrument development for the second study. Based on this study, preliminary conclusions are drawn about the linkages between evaluation policy and ECB.

Sample Selection for Study 1

Various organizations were selected for this study based on two criteria: (i) the selected aid or government organization had to have developed or revised its evaluation policy within the last five years; and (ii) the organization's website includes up-to-date information and documents about its evaluation policy and practices. Since I intended the evaluation policy search to be broad and far-reaching, I located evaluation policies through (1) searching the keywords evaluation policy, organization policy on evaluation, evaluation standards, evaluation criteria, evaluation guidelines, and evaluation strategy; (2) bibliographic follow-up from different reading done on evaluation policy; and (3) exchange with colleagues in a number of conferences. An initial list of the selected organizations was accumulated before starting the analysis (Appendix B). The sample includes 16 organization with old *and* revised policies and 20 organizations with a recently developed evaluation policy. Admittedly, I also considered documents with names other than evaluation policy but which include the categories of the evaluation policy, such as evaluation guidelines, evaluation strategy, evaluation manual, norms and standards for evaluation, results policy, and framework for evaluation. The decision to include such documents in the study was made after carefully reviewing each document according to the definition of policy provided above. All of the documents included in this study are official documents that have the elements of evaluation policy and are used to guide evaluation practice. As a measure of quality assurance, all of the documents came from the official website of the selected organization.

Study 1 Sample Characteristics

I conducted an extensive research and ultimately located 52 evaluation policies for 36 organizations as follows:

- 16 organizations with existing policy that had been revised (N= 32)
- 20 organizations with single evaluation policy (N = 20)
- Total N = 52 policies

As shown in Table 1, all 16 organizations included in the study with both old and updated evaluation policies reviewed their policies between 2012 and 2016. The majority of the organizations (11 of 19) with only one evaluation policy established their evaluation policy

between 2012 and 2014. Only eight of the organizations established their evaluation policy between 2009 and 2011.

Table 1 Evaluation Policies

Organization	Country	Context	Established Evaluation policy	Revised Evaluation policy
United States Agency for International Development (USAID)	United States	International development	2011	2016
World Intellectual Property Organization (WIPO)	Headquarters: Geneva, Switzerland	Global forum for intellectual property services, policy, information and cooperation	2010	2016
Treasury Board Policy on Evaluation (2009)/ Policy on Results (2016)	Canada	Government organization	2009	2016
United Nations Development Program (UNDP)	Headquarters: New York City, USA	International development	2006	2016
United Nations Evaluation Group (UNEG)	Headquarters: New York City, USA	Intergovernmental organization to prompt international cooperation	2005	2016
United Nations Industrial Development Organization (UNIDO)	Headquarters: Vienna, Austria	International organization to promote industrial development for poverty reduction, inclusive globalization, and environmental sustainability	2006	2015
United Nations Educational, Scientific and Cultural Organization (UNESCO)	Headquarters: Paris, France	International development	2007	2015
International Union for Conservation of Nature (IUCN)	Headquarters: Gland, Switzerland	International development; composed of both government and civil society organizations	2013	2015
Economic Commission for Latin America and the Caribbean (ECLAC)	Headquarters: Santiago, Chile	Regional economic and social development of Latin American countries (46 member states)	2005	2014
Swedish International Development Cooperation Agency (SIDA)	Headquarters: Stockholm, Sweden	Government agency: works on behalf of the Swedish parliament and government, with the mission to reduce poverty in the world	2004	2014
United Nations International Children's Emergency Fund (UNICEF)	Headquarters: New York City, USA	International development	2007	2013

Organization	Country	Context	Established Evaluation policy	Revised Evaluation policy
African Development Bank (AFDB)	Headquarters: Abidjan, Cote d'Ivoire	Regional multilateral development bank, engaged in promoting economic development and social progress in Africa	2007	2013
UK Department for International Development (DFID)	Headquarters: London, United Kingdom	Government aid organization	2009	2013
Canadian International Development Agency (CIDA)	Headquarters: Gatineau, Quebec, Canada	Government aid organization	2004	2013
Danish Development Cooperation	Headquarters: Copenhagen, Denmark	Government aid organization	2006	2012
World Health Organization (WHO)	Headquarters: Geneva, Switzerland	International development	2006	2012
International Labour Organization (ILO)	Headquarters: Geneva, Switzerland	International development: labour standards, social protection, and work opportunities	2005	2012
New Zealand Aid Program	Headquarters: Wellington, New Zealand	Government aid organization	NA	2014
United Nations Population Fund (UNFPA)	Headquarters: New York City, USA	International development	NA	2013
Australian Aid Agency	Headquarters: Canberra, Australia	Government aid organization	NA	2013
International Development Research Centre (IDRC)	Headquarters: Ottawa, Canada	Government aid organization	NA	2013
Spanish Agency for International Development Cooperation (AECID)	Headquarters: Madrid, Spain	Government aid organization	N/A	2013
French Development Agency (AFD)	Headquarters: Paris, France	Government aid organization	N/A	2013
Swiss Agency for Development and Cooperation (SDC)	Headquarters: Bern, Switzerland	Government aid organization	N/A	2013
Ministry of Foreign Affairs of Japan (MOFA)	Headquarters: Tokyo, Japan	Government organization	N/A	2012
United Nations Institute for Training and Research (UNITAR)	Headquarters: Geneva, Switzerland	International development	NA	2012
Millennium Development Goals Achievement Fund (MDG-F)	Headquarters: New York City, USA	International development	NA	2012

Organization	Country	Context	Established Evaluation policy	Revised Evaluation policy
Office for the Coordination of Humanitarian Affairs (OCHA)	Headquarters: New York City, USA	International aid organization: coordinates the global emergency response to save lives and protect people in humanitarian crisis	NA	2012
International Fund for Agricultural Development (IFAD)	Headquarters: Rome, Italy	International development	NA	2011
International Federation of Red Cross and Red Crescent Societies (IFRC)	Headquarters: Geneva, Switzerland	International aid organization	N/A	2011
United Nations High Commissioner for Refugees (UNHCR)	Headquarters: Geneva, Switzerland	International aid organization	NA	2010
Organisation for Economic Co-operation and Development (OECD)	Headquarters: Paris, France	Intergovernmental development organization	N/A	2010
Japan International Cooperation Agency (JICA)	Headquarters: Tokyo, Japan	Government development agency	NA	2010
Austrian Development Agency	Headquarters: Vienna, Austria	Government development agency	NA	2009
Dutch Ministry of Foreign Affairs	Headquarters: The Hague, Netherlands	Government organization	N/A	2009
United Nations Environment Program (UNEP)	Headquarters: Nairobi, Kenya	International development	N/A	2009

The organization contexts included in the sample depict evaluation policy in a number of international development organization and aid agency contexts, with some policies spanning different areas (e.g., Ministry of Foreign Affairs of Japan, 2012; Dutch Ministry of Foreign Affairs, 2009; Government of Canada Policy on Evaluation, 2009).

As shown in Table 2 below, the majority of the evaluation policies included in this study are United Nations (UN) organizations' policies. This is because the UN organizations' documents are available to the general public as a part of the UN commitment to accountability and transparency to country members and the public. The UN has an e-library called "The UN Documentation Research Guide" that allowed for searching the UN documents, publications, databases, and websites used for this study (<http://research.un.org/en/docs>). The sample of the evaluation policies spans the globe, with organizations located in North America, Europe, Africa, Australasia, East Asia, and Latin America and the Caribbean.

Table 2 Distribution of Evaluation Policies by Geographical Region

Region	Organization
Africa (N=1)	• AFDB: African Development Bank
Australasia (N=3)	• Australian Aid Agency • Australian Development Agency • New Zealand Aid Program
North America (N=4)	• CIDA: Canadian International Development Agency • Treasury Board of Canada • IDRC: The International Development Research Centre • USAID
East Asia (N=2)	• JICA: Japan International Cooperation Agency • MOFA: Ministry of Foreign Affairs of Japan
Europe/ UK (N=8)	• Danish Development Cooperation • Spanish Agency for International Development Cooperation • French Development Agency • Dutch Ministry of Foreign Affairs • SDC: Swiss Agency for Development and Cooperation • IUCN: World Conservation Union • DFID: UK Department for International Development • SIDA: Swedish International Development Cooperation Agency
UN (N=15)	• ILO: International Labour Organization • UNICEF: United Nations International Children's Emergency Fund • UNIDO: United Nations Industrial Development Organization • WIPO: World Intellectual Property Organization • UNESCO: United Nations Educational, Scientific and Cultural Organization • WHO: World Health Organization • UNDP: United Nations Development Program • UNEG: United Nations Evaluation Group • IFAD: International Fund for Agricultural Development • UNITAR: United Nations Institute for Training and Research • UNFPA: United Nations Population Fund • MDG-F: Millennium Development Goals Achievement Fund • OCHA: Office for the Coordination of Humanitarian Affairs • UNEP: United Nations Environment Program • UNHCR: United Nations High Commissioner for Refugees
International (N=2)	• IFRC: International Federation of Red Cross and Red Crescent Societies • OECD: Organisation for Economic Co-operation and Development
Latin America and the Caribbean (N=1)	• ECLAC: Economic Commission for Latin America and the Caribbean

Study 1 Review Strategy and Analysis

After identifying the 52 evaluation policies selected for this study, I read each policy closely several times in order to acquire a sense of the overall purposes and components of the organizations' evaluative practices and to ascertain what I believed to be meaningful to my study within the data. Throughout this process, I summarized each evaluation policy and, over time, I added the summaries to a descriptive summary table (Appendix C) noting the

name of the document, the year of publication, the context, the purpose of the policy and the main focus. The process of summarizing the policies enabled a greater understanding of the content of the policies and facilitated the analysis.

The analytic strategy used in this study is guided by the same conceptual framework that formed the basis of the development of the research questions, the research design, and the data collection methods that I selected for this study. I analyzed the data using NVivo software. I input evaluation policy documents into the software as PDF sources and applied the codes directly to the evaluation policies themselves.

The original framework for the analysis consisted of Trochim's (2009) eight categories of evaluation policy, which I used as a primary coding list to analyze the data. The eight categories are: (1) evaluation goals; (2) evaluation participation; (3) evaluation capacity; (4) evaluation management; (5) evaluation processes and methodology; (6) evaluation roles; (7) evaluation uses; and (8) meta-evaluation. It should be noted that Cooksy, Mark and Trochim (2009) call attention to the need for empirical validation of these eight categories of evaluation policy. They suggest that these categories be checked against existing evaluation policies to see the extent to which they are comprehensive and represent categories of actual evaluation policies. Therefore, these eight categories served as the basis for the subsequent policy analysis. As the coding and systematic analysis of the evaluation policies proceeded, new categories emerged, and I integrated and separated some categories as the data took shape. Having identified emerging categories, I used the constant-comparative method (Cresswell, 1998) to refine the initial categories and validate my preliminary findings. This process involved multiple steps as the data under each category were analyzed separately to ensure that they were relevant and representative of that category. In the cases where the data did not fit any of the eight categories, a new category was created. The categories were compared for overlap and were refined throughout the process.

Ultimately, I was able to identify 10 broad categories that capture the goals, processes, strategies, criteria and activities of evaluation policy, of which eight are Trochim's (2009) categories (albeit with some adjustment and integration) and two are additional categories that stood out during the analysis as being employed in a number of policies. The 10 broad categories are: (1) evaluation goal; (2) evaluation definition; (3) evaluation role; (4)

evaluation management; (5) evaluation norms and standards; (6) evaluation planning and processes; (7) evaluation use; (8) meta-evaluation; (9) ECB; and (10) competencies and ethics. I believe these categories to be sufficiently unique to merit specific attention.

The main results of the substantive analyses are reported in Chapter 4. Chapter 4 is structured according to the conceptual framework. My general approach was to look across the evaluation policies and provide an integrated explication of the main findings within that structure. With such a large sample ($N = 52$), I did take great care to associate a good representation of the evaluation policies with claims that emerged from my analysis. I also provide illustrative verbatim excerpts and examples that support the claims I made.

Study 2: Interviews with Evaluation Scholars and Practitioners (Research Questions 1, 2, & 3)

This study consists of interviews with evaluation scholars and practitioners who have contributed substantial published work in the areas of ECB and/or evaluation policy. Given that a significant literature on evaluation policy does not exist, I decided to interview scholars who have experience writing and theorizing about evaluation policy and/or organizational ECB, as well as individuals who have practical experience working in the field, as a way to develop (i) a more focused and in-depth understanding of evaluation policy and its role in building organizational capacity for evaluation and (ii) deeper first-hand interpretations both of the role of evaluation policy in building organizational evaluation capacity and of the multiple and varied dimensions at play within an organizational context. Essentially I followed the lead of other researchers in the field of evaluation, such as Christie (2003) and Chouinard (2010), who interviewed evaluation scholars in their doctoral work as a way to foster dialogue among evaluators and, ultimately, to inform the theory and practice of evaluation. The guiding questions for Study 2 are set out in Appendix D.

Study 2 Sample Selection and Characteristics

I selected the participants in this study primarily because they have published in the field and/or have significant experience working on organizational capacity building and evaluation policy. I identified potential participants in a number of ways: through their previous publications in the field, through their active involvement in relevant special interest

groups such as American Evaluation Association Topical Interest Groups (Evaluation Policy, ECB, Evaluation Use and Influence) and collegial recommendations. In total, I identified 22 potential participants and developed a list noting their names and email addresses as well as the rationale for their recruitment for my study. My thesis supervisor, Professor Cousins, initiated the communication with the potential participants by sending an introductory email to explain the nature and objectives of the research (Appendix E). Once each person had agreed, I followed up by sending an official invitation providing additional details about the study and about what participation would entail (Appendix F). Attached to this email was an individual letter of informed consent (Appendix G).

Ultimately, 18 evaluation scholars and practitioners from Canada, the United States, and Europe agreed to participate in a telephone interview. While the majority of the participants are university professors, some of them are working as evaluation consultants and/or with government or private organizations. The participants have between 15 and 30 years of experience working as evaluators or in evaluation contexts, both nationally and internationally.

Study 2 Data Collection

Study 2 consists of 18 one-on-one semi-structured interviews with evaluation scholars and practitioners with the objective of obtaining their views on how evaluation policy influences organizational evaluation capacity and why this influence occurs. The interviews were audio-recorded with the permission of the participants and were subsequently transcribed by me verbatim in Word. I also took notes during the interviews to add personal observations about the participant or flag particular comments for review.

The interviews were based on in-depth, semi-structured, and open-ended questions and lasted between 60 and 90 minutes. I based the interview protocol (Appendix D) on issues and outstanding questions identified through the comprehensive review of evaluation policies completed in Study 1. The first question focused on setting the stage for the study and the remainder of the interview and asked the participants to define evaluation policy and organizational capacity to do and use evaluation. The subsequent questions were designed to help the participants reflect on evaluation policy by situating them in an organizational setting and inviting them to provide information and opinions based on the concrete

manifestations of evaluation policy in relation to evaluation practice (e.g., What impact do you think evaluation policy has on organizational capacity for evaluation?).

As a measure of quality assurance, I pilot tested the interview guide prior to the start of data collection. I asked two colleagues, both professional evaluators, to review and comment on the interview guide in terms of the clarity of the questions and the content and comprehensiveness of the instrument. Based on their feedback, I made only minor changes to the wording to improve clarity. The changes did not affect the overall sense of the questions included in the guide.

Study 2 Data Analysis

I used qualitative data analysis to identify recurring themes and patterns in the interview data. As mentioned, I transcribed all 18 of the interviews into individual Word files. The Word files were then integrated into NVivo as sources, and I applied codes directly to them. Rich data has to do with the quality of the data collected, as well as with the use of multiple corroborating data sources (Maxwell, 2005). This study consisted of more than 18 hour-long interviews with scholars and practitioners that were subsequently transcribed verbatim to allow for more in-depth analysis.

The original framework for the analysis of the interviews consisted of Bourgeois and Cousins' (2008, 2013) six dimensions of organizational-evaluation capacity: (1) human resources; (2) organizational resources; (3) evaluation planning and activities; (4) evaluation literacy; (5) organization decision-making; and (6) learning benefits. The analytic strategy I employed was to use these dimensions as an initial list of codes while remaining open to the possibility of identifying additional emergent codes. This strategy allowed me to identify themes and categories associated with the ways in which the participants understand evaluation policy and its role in organizational capacity for evaluation. The analysis comprised two stages. During the first stage, I focused on reading through each transcript, taking notes to flag specific pieces of information, and identifying any recurring themes or emergent categories. As a result of this process, I identified a significant number of emergent codes/categories across all of the interview data. This led me to explore the relationships between and within these codes/categories in the second stage of the analysis. The second stage involved the classification of the categories according to their level and focus with

regard to the relationship between evaluation policy and organizational capacity for evaluation. Guided by the research questions and the findings from the comprehensive policy review in Study 1, the analysis of the interview data ultimately led to my development of a conceptual framework as a way to sort and make sense of the identified themes and categories.

An explanation of the elements of the framework and the findings of this study are detailed in Chapter 5 of the thesis.

Study 3: Focus Groups with MENA Evaluation Community Members

This study consists of two focus group discussions with evaluation community members and evaluation professionals in the Middle East and North Africa region (MENA) and specifically from two countries, Jordan and Turkey. This study aims to gain an understanding of the extent to which evaluation policy plays a role in the region and of the connections that might be made with organizational evaluation capacity. MENA was selected for this study for a number of reasons:

- (i) While we already have some knowledge of evaluation policy in the West where evaluation has a relatively long history, we currently know little about evaluation policy in the MENA region.
- (ii) By focusing on MENA, a region where evaluation capacity and practice are relatively underdeveloped, it is likely that we will gain new insight into the links between evaluation policy and capacity building.
- (iii) There is a growing interest in evaluation in the region, as is reflected in the establishment of the Middle East and North Africa Evaluators Network (EvalMENA) in 2011 and a recent conference held in Istanbul in February 2015 as part of the 2015 the International Year of Evaluation. Also, the two countries that I selected recently established voluntary organizations for professional evaluation (VOPEs), specifically the Jordanian Development Evaluation Association (EvalJordan) in 2014 and the Turkish Monitoring and Evaluation Society (TMES) in 2013.
- (iv) I am from the region, familiar with the cultures, and committed to making a contribution to the region that will benefit the entire evaluation community.

While the interviews in Study 2 highlighted aspects of evaluation policy from the perspective of the Western evaluation scholars and practitioners, the contextual dimension was only incidentally present (i.e., there was some variation over context within the confines of Western societies)—specifically the perspective of the evaluation community members on evaluation policy and how it influences their practice. The purpose of the focus groups was to discuss the findings of Study 2 with the members of the MENA evaluation community and explore the extent to which they resonated in this distinct cultural context. I was interested in ascertaining the extent to which they aligned with the experiences of evaluators and organization members working in the MENA region and the ways in which evaluation policy influence organizational evaluation capacity and evaluation practice in developing- and transitional-country contexts. The following questions guided the two focus groups:

- To what extent do organization evaluation policies exist in the region?
- Does evaluation policy play a role in fostering organizational capacity for evaluation? To do evaluation? Or to use it?
- If yes, what conditions or variables need to be in place?

These questions map onto overarching research question 3; they provide a direct opportunity to explore contextual issues whereas in Study 2 such issues were somewhat sporadic and incidental.

Study 3 Sample Selection and Characteristics

I selected two evaluation communities in the MENA region to participate in this study—EvalJordan and TMES—because of their efforts and interest in building evaluation capacity within their respective contexts. Arrangements were made to conduct the two focus groups with the evaluation communities in their home countries of Turkey and Jordan. Ultimately, I travelled to Turkey and Jordan with as research associate of my supervisor Professor Brad Cousins, who accompanied me to assist in facilitating the focus-group discussions. I provide further information about each of the focus groups below.

Focus group 1: The Turkish Monitoring and Evaluation Society (TMES). TMES was established in 2013 but had become reenergized in October 2017, only two months before the focus group was conducted. This provided an excellent opportunity to discuss TMES members’ perspectives on evaluation policy and the ways in which these perspectives

could influence their respective practices during the early stages of the society's development. The discussions acknowledged and were influenced by the fact that TMES aims to foster an evaluation culture in Turkey by strengthening both the supply side and the demand side of evaluation and by undertaking a series of capacity-building activities within the country. It should be noted that many consider Turkey to be an economically transitional country and therefore its reliance on international aid is quite limited, particularly in comparison with developing countries.

The focus group was held in Ankara, Turkey with eight participants. The participants were identified and recruited using snowball sampling (Rubin & Babbie, 2001), with the assistance of a previously identified community member. The participants were invited because of their active role in Turkey's evaluation community. While the majority of the participants were working within government organizations, some were university professors. The meeting was scheduled late on a day when members of all stakeholder groups in the evaluation community in Turkey were having a meeting concurrently in a conference organized by the Ministry of Science, Technology and Industry. Some participants were from different regions of Turkey although most were located in Ankara, the nation's capital. Professor Cousins initiated the communication with and the invitation to the members of the focus group via email, and I followed up with an email outlining my research and defining the parameters of involvement in the focus group. The email also included (i) a document providing an overview of the research and a summary of my preliminary findings from Study 2 for the participants to read in preparation for the focus group (Appendix H); and (ii) an individual letter of informed consent, of which signed versions were to be collected before the focus group (Appendix I).

Focus group 2: Jordanian Development Evaluation Association (EvalJordan).

EvalJordan was established in 2014 with the goal of improving evaluation practice in Jordan and supporting the government and civil society organizations in making informed decisions about policies and programs. I selected EvalJordan to participate in this study because of its active role in promoting evaluation practice, not only in Jordan but also in the MENA region as a whole, and because of its continuous investment in capacity-building activities. The association is supported and legally recognized by the Jordanian government

to cultivate an evaluation culture in the country and create a platform for evaluators and institutions that are in need of high-quality evaluations.

Jordan is a developing country in which multilateral and bilateral aid organizations are active, and the country's capital, Amman, is the location of the regional office of a number of aid and development agencies (e.g., the UNICEF regional office). The focus group took place in Amman with 10 participants. All are members of EvalJordan and, as was the case with the TMES focus group, were identified and recruited using snowball sampling. The participants are employed by government, aid, and private organizations. A colleague serving as president of EvalJordan and chair of the umbrella organization EvalMENA facilitated the communication with association members to invite them to participate in this study. Once they agreed to participate, I followed up with an email describing my study and providing details about what participation would entail. The email also included (i) a document providing an overview of the research and a summary of my preliminary findings from Study 2 for the participants to read in preparation for the focus group (Appendix H); and (ii) an individual letter of informed consent (Appendix I).

Study 3 Data Collection

The focus groups were designed to obtain a better understanding of the MENA evaluation community members' perspectives on evaluation policy in relation to capacity building within their respective contexts. The focus groups also facilitated a discussion with the participants about the three-level framework that I had previously developed in Study 2 of my research and about its applicability to their particular contexts.

The focus groups were based on open-ended questions with the total of 18 participants, and they lasted for about 90 minutes. The guiding questions focused on issues related to the MENA context and the extent to which evaluation policy plays a role in fostering organizational capacity within this context (Appendix I). The first question focused on setting the stage for the focus group and asked the participants about the extent to which organization evaluation policies exist in the region. Various questions were designed to help the participants reflect on the role played by evaluation policy in fostering evaluation capacity (e.g., Does evaluation policy play a role in fostering organizational capacity for evaluation? To do evaluation? Or to use it?) and on the contextual factors that influence this

role (e.g., What conditions or variables need to be in place for evaluation policy to foster capacity?).

The focus groups were audio-recorded with the permission of the participants. Professor Cousins facilitated both of the focus groups and provided me with his observations and impressions of the discussions. Also, as a measure of quality assurance, Professor Cousins and I revised and streamlined the focus group guide prior to the meeting in Ankara. We also debriefed after the focus groups and shared our impressions.

Study 3 Data Analysis

As with Study 2, I used qualitative content analysis to identify themes and trends in the focus group data. I summarized the audio recordings of the two focus groups into individual Word files taking care to identify specific illustrative verbatim quotations to use them as examples when I reported the findings. The Word files were then integrated into NVivo as sources and codes were applied directly to them. Before selectively transcribing the two focus groups, I listened to each audio recording twice to develop a sense of the flow of the discussion and the content for greater clarity and understanding. As I listened, I took notes on impressions, hunches, and possible themes, looking for convergence across the two focus groups (Patton, 2002). When I coded the data, I divided the individual pieces of information into comments about evaluation policy and capacity building within the context of the country on the one hand and comments on the components of the three-level framework on the other in order to clearly distinguish between the two groups of data and identify areas of overlap between the two focus groups. The findings of this study are located in Chapter 6 of the thesis.

Ensuring Data Quality

In undertaking this qualitative study, I relied upon several strategies to address the trustworthiness and quality of the data collection and the interpretation of my results. First and foremost, as described above, I used a comprehensive cascading research design which draws from conceptual, methodological and substantive domains. The three studies are logically sequenced and collectively permit the comprehensive coverage of the questions guiding the research. Also described above, I defined and described my own epistemological

stance and demonstrated how the research design that I selected was consistent with it. Other important strategies were triangulation of methods and data sources, collection of rich data, peer review and creation of an audit trail. In what follows, I provide a brief description of each of the strategies adopted in order to strengthen the quality of my findings.

Triangulation refers to the use of multiple methods and data sources to develop a comprehensive understanding of the research issues and reduce the risk of systemic biases emerging as a result of the use of a specific method (Creswell, 1998; Maxwell, 2005; Patton, 1999). In the three research studies, I used multiple methods to collect my data, including a comprehensive review of evaluation policies, interviews, and focus groups. I also collected data from a range of individuals, including evaluation scholars and practitioners as well as evaluation community members from different countries. Further, the participants in Study 3 were asked to respond specifically to the conceptual framework arising from Study 2, thus helping to provide validation of the prior findings.

Rich data refers to the quality of the data collected and to the use of multiple corroborating data sources (Maxwell, 2005). My study includes rich and thick verbatim descriptions of participants' accounts in order to support the findings. The 18 one-hours interviews with evaluation scholars and practitioners that I subsequently transcribed verbatim allowed for in-depth analysis. I followed up the interviews with two separate focus groups, located in two distinct Middle Eastern contexts, held with 18 evaluation community members as a way to test my findings and enhance their credibility and trustworthiness.

Peer review involves an external check on the researcher to play devil's advocate, challenge the researcher's assumptions, check for potential researcher biases, and clarify interpretations (Creswell, 1998; Lincoln & Guba, 1985). I used this strategy periodically over the course of my entire research process to help increase the quality and trustworthiness of my findings. For example, I made numerous presentations, both formal and informal, to colleagues from the University of Ottawa and at various conferences. I presented the conceptual framework for discussion at a number of conferences in Europe, the United States, and Canada (e.g., the European Evaluation Society Biennial Conference in Dublin, 2014; the American Evaluation Association Conference in Chicago, 2014; and the Canadian Evaluation Association Conference in Montreal, 2015), sharing my initial conceptual framework and data collection strategy. The framework was well received by those in

attendance and generated fruitful discussions that helped to clarify my conceptual framework and its associated components.

Moreover, I pilot tested the interview guide for Study 2 prior to the start of data collection. I asked two colleagues, both professional evaluators, to review and comment on the interview guide in terms of the clarity of the questions and the content and comprehensiveness of the instrument. Based on their feedback, I made only minor changes to the wording to improve clarity. These changes did not affect the overall sense of the questions included in the guide.

I also engaged my supervisor during Study 3 of my research to facilitate and observe my focus groups, and to provide me with his impressions about what had transpired and what the participants had said. After each focus group, we debriefed and discussed in detail the participants' responses to my questions, and we shared our reflections about the content of the participant discussions.

Audit trail involves keeping a research log about how decisions are made throughout the study, primarily regarding the selection of methods and the analysis of data (Guba & Lincoln, 1985; Merriam, 2002). To capture the decisions made over the course of my research, I kept a detailed log of my ideas about data collection and analysis, of conversations with colleagues that influenced changes in my thinking, and of the process decisions that I made along the way. I also captured my impressions and thoughts immediately after each interview and focus group. This helped me to reflect on my impressions and understanding and to plan the next steps of my research.

Chapter 4: Evaluation Policy Analysis Findings

This chapter presents the findings obtained through the comprehensive and critical review and analysis of a sample of 52 evaluation policies of various multilateral and bilateral aid agencies and government organizations that have recently developed or revised their evaluation policy. The purpose of this study was to scan the landscape and to understand and elucidate the current trends in evaluation policy development and their connection with evaluation capacity. More specifically, the purpose was to identify the components of evaluation policy that influence organizational evaluation capacity in order to answer Research Question 1 on the role of evaluation policy in the organizational evaluation capacity and the extent to which development of organizational evaluation capacity is espoused in existing organizational evaluation policies.

The following two questions provided an initial focus to guide the review:

1. What rationale is given for the existence/development of the evaluation policy of different organizations?
2. What categories are included in the various evaluation policies?

Of particular interest is the extent to which elements of evaluation policy relate to organizational capacity to do and use evaluation.

The analysis is guided by the conceptual framework (see Figure 1) in order to identify and examine the categories of evaluation policy. I subsequently used these findings to inform instrument development for Study 2. Based on this review, I drew preliminary conclusions about the linkages between evaluation policy and ECB. One of the main outcomes of this study is the identification of 10 evaluation policy categories that cover various aspects of the evaluation function within the organization. Also, a significant outcome of this study is the articulation of a definition of evaluation policy as a result of a closer examination of the elements of evaluation policy as reflected in the policy documents.

Comprehensive Review and Analysis of Evaluation Policies

Through a comprehensive review and analysis of an extensive sample of evaluation policies, I was able to identify 10 broad categories that capture the goals, processes, strategies, criteria, and activities of evaluation policy, of which eight are Trochim's (2009)

categories (albeit with some adjustment and integration) and two are additional categories that stood out during the analysis as being employed in a number of policies.

The 10 broad categories are: (1) evaluation goal; (2) evaluation definition; (3) evaluation role; (4) evaluation management; (5) evaluation norms and standards; (6) evaluation planning and processes; (7) evaluation use; (8) meta-evaluation; (9) ECB; and (10) competencies and ethics. I believe these categories to be sufficiently unique so as to merit specific attention.

Evaluation Goal

This category of evaluation policy focuses on making understandings related to the purpose of the organization's evaluation function explicit and transparent for all stakeholders. All 52 of the policies included in this study identified the goal of evaluation within their respective organizations. In most policies, the goal of evaluation is stated in the first section under such titles as: (1) the purpose of evaluation; (2) fundamental purposes of evaluation; (3) the role of evaluation; (4) goal of evaluation; (5) why we carry out evaluation; (6) evaluation objectives; (7) rationale for evaluation; (8) purpose and use of evaluation; (9) why we need high quality evaluation; (10) evaluation aims and scope; and (11) goals, strategic direction and the purpose of the evaluation function. In some policies, the evaluation goals are included in the introduction, background, or context sections of the document.

Accountability, learning, and developing/supporting an 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 reason why organizations conduct evaluation. Accountability is also a major reason why many organizations established their evaluation policy in the first place. The data reveal that accountability, as a major goal for evaluation, is linked to evidence-based practices and performance measurement within the organizations. Here are some excerpts by way of illustration.

The primary purpose for conducting evaluations is to provide substantive accountability for the resources provided to the organization to implement its programme activities and the results from the implementation of these programmes. For that reason, accountability is a core principle on which evaluations are based. (UNDP, 2016, p. 9)

Evaluation has two main purposes: to help managers design or improve the design of policies, programs and initiatives; and to provide periodic assessments of policy or program effectiveness, of impacts both intended and unintended, and of alternative ways of achieving expected results. (Treasury Board Policy on Evaluation, 2009, p. 2)

The policy seeks to increase transparency, coherence and efficiency in generating and using evaluative knowledge for organizational learning and effective management for results, and to support accountability. (UNDP, 2011, p. 2)

Evaluation serves two interrelated main purposes. First, 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. Second, evaluation is a means to facilitate learning about the positive and negative experiences of development cooperation in specific contexts. (Danish development cooperation, 2012, p. 7)

[Evaluation] ensures accountability and oversight for performance and results, and reinforces organizational learning in order to inform policy for decision-makers and support individual learning. (WHO, 2012, p. 1)

Evaluations help hold UNDP and its associated programmes and funds accountable to stakeholders. Evaluations thus constitute an important source of evidence for monitoring organizational performance. (UNDP, 2016, p. 2)

IFAD recognizes that one main purpose of evaluation is to promote accountability. Accountability in this context refers to the assessment of developmental results, the impact of development assistance and the performance of the partners involved. (IFAD, 2011, p. 7)

Thus, evaluation policies' goal of accountability is focused on improving the internal practice of evaluation within the organization as well as the levels of accuracy and transparency in responding to external demands, such as demonstrating accountability for the use of public funds. By identifying accountability as a goal for evaluation, organizations acknowledge their own responsibility for their decisions, roles, and actions, as well as for the products of evaluation.

Learning is the other goal of evaluation that is explicitly identified in almost half of the evaluation policies. To a large extent, the findings reveal that focusing on accountability as a goal for evaluation is perceived as fostering organizational learning from the evaluation. The following is an illustration:

Accountability is considered as a necessary first step in the learning process. Systematic independent evaluation of completed projects and past and ongoing policies and strategies are indispensable if IFAD is to learn from its experience, both positive and negative, and improve its future effectiveness. (IFAD, 2011, p. 9)

All evaluations share the dual purpose of organizational learning and accountability. (UNESCO, 2015, p. 4)

Learning is stated as a principal goal for evaluation in all of the recently revised evaluation policies and in some of the newly developed policies. This focus on learning reflects many organizations' commitment to evaluation as a system to increase knowledge and improve performance levels on an ongoing basis. This in turn promotes continuous improvement and supports the achievement of organizational goals and the ability to deal with changing contexts, as implied in several policies as shown in 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. (Austrian Development Agency, 2013, p. 4)

Evaluations in SDC promote organisational learning for strategic guidance and better programme management. They are indispensable instruments for strengthening the results orientation in complex and often fragile contexts. (SDC, 2013, p. 5)

Evaluations can systematically generate knowledge about the magnitude and determinants of performance, permitting those who design and implement them to refine designs and introduce improvements into future efforts. Learning requires: careful selection of evaluation questions to test fundamental assumptions underlying strategies and project designs; methods that generate findings that are internally and externally valid; and systems to share findings widely and facilitate integration of the evaluation conclusions to recommendations into decision-making. (USAID, 2016, p. 2)

Many organizations thus explicitly recognize that learning from evaluation must be included as a goal in evaluation policies in order to ensure that evaluation practices and activities are strategically focused and goal-directed to generate relevant knowledge. Indeed, according to the UNDP policy,

Evaluations serve course-corrective decision-making by way of evidential data collection, reflection and analysis, as well as impartial judgement. The scope, design and implementation of an evaluation should generate relevant, cost-effective and timely information. All evaluations should drive organizational learning. (UNDP, 2016, p. 2)

A number of evaluation policies also note that the goals of evaluation extend beyond accountability and learning to 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)

IDRC strives to be an accountable learning organization by integrating a culture of evaluative thinking into its activities . . . evaluation must go beyond conduct and promulgation of formal evaluation studies. (IDRC, 2013, p. 1)

UNHCR recognizes the need to develop an organizational culture that is amenable to the development of an effective evaluation function. In that respect, the Office is striving to ensure that managers and staff members are fully committed to evaluation and learning. (UNHCR, 2010, p. 1)

As such, evaluation serves the purpose of promoting accountability and learning and forms the basis for fostering evaluative culture within the organization. The analysis reveals that evaluation policies set the broader framework for evidence-based practices and performance measurement systems whereby learning and accountability converge and accountability is taken to mean taking responsibility and being committed to evaluation goals, roles, and strategies. Analysis of this category indicates that there seems to be some interest in conceptualizing evaluation as a learning function, and not just in accountability, which is likely to be favourable to ECB.

Evaluation Definition

The majority of the evaluation policies include a clear definition of the evaluation concept and what it entails within the organization. This is the second most frequently mentioned category of evaluation policy—an additional category that was not included in the original framework but instead emerged from the data. As stated in Chapter 2 of this thesis, the fact that evaluation is a *transdiscipline* that encompasses diverse theories, methods and practices contributed to the difficulty encountered in working to develop a unified definition of evaluation. The data reveal that 44 evaluation policies included an explicit definition of evaluation to clarify the concept and to distinguish between evaluation and other closely related concepts and functions within organizations, such as audit, inspection, and measurement. The majority of the evaluation policies started with a definition of evaluation under such headings as: (1) evaluation definition; (2) what is evaluation; (3) definition of the evaluation concept; and (4) concepts and consistent terminology. In some cases, the definition of evaluation is included in the introduction to the policy document.

The crucial importance of the definition of the concept of evaluation is largely self-evident. The inclusion of an explicit and contextual definition of evaluation in the evaluation policy is essential for ensuring a common understanding and eliminating any ambiguity

about the meaning of the evaluation function for the organization. The data reveal that many policies include a statement not only of what evaluation is but also of what it is not. An example of such a statement is as 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. Evaluation focuses on results, design, implementation and management processes. It should not be confused with implementation monitoring and reporting, audit, inspection, investigation or assessment of individual performance. Although it takes the form of data based analysis, evaluation is not academic research. (ILO, 2005, p. 2)

Since evaluation policy is a type of communication tool, many organizations acknowledge that the communication of what they mean by evaluation is both necessary and important. Without an explicit definition in the policy document, the meaning of evaluation is open to a wide range of assumptions and interpretations. In the data there are many specific examples where the definition of evaluation includes a distinction between the evaluation function and the other functions in the organization, as in the following example:

The distinction between performance audit and evaluation can be thin as typically both deal with the “what” and “why” questions of programme and organisational performance.

A useful way to distinguish performance audit from evaluation generally is to highlight that performance audit is carried out primarily to provide assurance, i.e. accountability, on a direct or attest basis, with knowledge generation a secondary spin-off. Evaluation on the other hand is carried out usually to inform policy and decision-making as well as to produce knowledge. (Danish Development Cooperation, 2006, p. 11)

While a number of policies include definitions of specific types of evaluation (e.g., formative evaluation, impact evaluation), other policies include a glossary of key terms used when defining evaluation and/or its standards and criteria (see, for example, the evaluation policies of UNESCO, IUCN, and IFAD). Thus, evaluation definition is an important category in the evaluation policy because it captures the meaning, function, and use of evaluation within the organization. This is likely to be favourable to ECB because there is clarity about what evaluation is and is not and about what it can be expected to accomplish.

Evaluation Role

This category is concerned with 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 its responsibilities and

accountabilities among individuals who have been entrusted with multiple distinctive roles. The vast majority of the evaluation policies include a specific section about the roles of and responsibilities for the evaluation function within the organization. This category of the evaluation policy is explicitly included in 43 evaluation policies and was identified in the framework for evaluation policy detailed by Trochim (2009).

The data indicate that clearly defined roles and responsibilities are important for carrying the evaluation function so that every individual who is involved in the evaluation is clear about his or her responsibilities, including reporting and accountability roles. Specifically, the evaluation policies identify job positions and include detailed information about their respective responsibilities and expectations, about the executives, managers and/or other stakeholders to whom they report, and about the level of authority required when making decisions about the evaluation. This category is particularly important for ensuring that no part of the evaluation workload is overlooked and for maintaining effective management of the evaluation processes. Some examples are as follows:

(a) The General Conference and Executive Board collectively safeguard the independence of the evaluation function. . . . The Executive Board discusses evaluation findings and takes appropriate action to inform organizational policy, strategy and programs. (b) The Director-General is accountable for evaluation results. S/he assures the integrity and independence of the evaluation function and creates an enabling environment, which recognizes the importance of evaluation as a key accountability and learning mechanism. (UNESCO, 2015, p. 5)

The Chief Controller reports to the Director-General every six months about the results of the follow-up of the implementation of action plans made in response to evaluations undertaken by UTV. The Chief Controller should also keep UTV informed about the results of this monitoring process. UTV shall regularly review the extent to which measures are taken as a result of evaluation findings, recommendations, and action plans. Annually, a comprehensive analysis of the findings and lessons learnt of all Sida evaluations should be carried out for which the Chief Controller is responsible in co-operation with UTV and Sida departments. (SIDA, 2014, p. 21)

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

Well-articulated evaluation policies lead to more efficient and effective evaluation function. This is because people function most efficiently when they share a common understanding of

the respective roles and responsibilities of the various stakeholders. It is argued that one of the reasons why evaluations sometimes fail is because of the lack of clarity among the organization's members regarding their respective roles and responsibilities for the evaluation. When roles and responsibilities are clearly defined in the evaluation policy, there is less duplication of effort, less confusion, greater productivity, and better use of results. The data show that there are many examples where the use of evaluation results is incorporated into the roles and responsibilities of identified members within the organization. For example:

IUCN senior management, which includes the Director General, the Deputy Director General, Regional Directors and Global Directors, 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. Management is expected to be an active participant in M&E activities in terms of collaboratively setting a work plan, participating in design, considering draft reports and using results in decision making. Senior management also monitors the implementation of changes as recommended by an evaluation. (IUCN, 2015, p. 12)

Some evaluation policies also include specific roles and responsibilities for implementing the policy itself. Because the implementation process has the potential to impact a wide range of duties and areas of responsibility, organization members need to have a thorough understanding of their individual and collective responsibilities throughout the implementation process and thereafter. For example:

The Evaluation Department leads and facilitates the implementation of the Evaluation Strategy through:

- Coordinating the process for determining evaluation priorities within each thematic policy area jointly with the relevant policy and RED teams
- Proposing the initial shortlist of high priority evaluations that will receive enhanced central support and providing co-management of the approved high priority evaluations
- Supporting policy teams in identifying and co-managing thematic evaluations
- Supporting quality improvement processes
- Designing and implementing an evaluation communications plan
- Producing and Annual Evaluation Report
- Monitoring the implementation of the Evaluation Plan and reporting annually to the Investment Committee. (DFID, 2013, p. 26)

This category of the policy helps organization members to look beyond their own position and to understand the roles and contributions of the other members. It is also likely to be favourable to ECB because it recognizes the fact that the overall success of the evaluation is a function of the degree to which responsibility and ownership are shared within the organization with a view to achieving the evaluation goals and attending to the organization's objectives/mission/vision.

Evaluation Management

This category of evaluation policy focuses on important aspects of the management of the evaluation function within the organization, including evaluation budgeting and resource allocation, evaluation-related human resources management and evaluation timing. This category was included in Trochim's (2009) framework for evaluation policy and has been recognized by various types of organizations included in the study.

An effective evaluation function requires a secure and adequate investment into financial and human resources management. This investment is necessary to ensure the development of a professional evaluation function that produces the evidence and insights required for decision-making and that creates an evaluative culture that fosters organizational learning. The data reveal that this category is evidenced in most of the evaluation policies and is explicitly recognized in 39 of them. Most of the policies include sections on: (1) evaluation management; (2) evaluation budget; (3) resources for evaluation; and (4) human resources management. The following provides an example of the evaluation management category:

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. This will mean that adequate lead-time must be allocated to design and commission the evaluation. Each mission must conduct at least one "whole-of-project" performance evaluation within their timeframe. Such evaluations are particularly valuable in the period preceding the development of a new strategy, when questions are likely to be asked about the overall effectiveness of engagement in a particular sector or broad set of activities. (USAID, 2016, p. 9)

The majority of the policies included in the study sample also provide details about the importance of the evaluation budget when managing the evaluation. The data show that many organizations are explicit about the evaluation budget, with some indicating the specific

amount of money directed to the evaluation function. The majority of the policies state that the evaluation budget is between 3 and 5 percent of the entire project budget. Some examples of these kinds of budget allocations are as follows:

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)

The costs of an evaluation must be in an appropriate proportion to the scope of the project. In general terms, approximately 3–5% of the entire project or programme budget is provided for an evaluation. (Austrian Development Agency, 2009, p. 5)

UNFPA will allocate funds for the Evaluation Office and the evaluations using separate budget lines in the UNFPA integrated budget. It will allocate up to 3 per cent of the total program budget for the evaluation function. The Evaluation Office will manage the budget for the staffing and operational costs of the office. (UNFPA, 2013, p. 13)

In determining the amount required to finance evaluation in WHO, estimations provided by other organizations have been considered. According to these, the overall programme budget might contain, as an integral part, a figure for evaluation that is equivalent to between 3% and 5% of that budget. (WHO, 2012, p. 7)

Some evaluation policies go beyond specifying the evaluation budget to articulate guidelines for planning and calculating evaluation costs. Such guidelines are intended to help manage, analyze and evaluate financial performance, as can be seen in the following illustration:

In order to calculate the evaluation budget for financing review missions, these six steps should be followed: 1. Calculation of number of consultant working days. 2. Determination of consultant level according to expertise and experience. 3. Calculation of travel costs, including travel days, vehicle use for field trips. 4. Calculation for data collection, either primary or secondary. 5. Calculation of accommodation and DSA costs. 6. Calculation of any additional costs (interpretation services, workshop facilities for focus group and stakeholder meetings, etc.). (ILO, 2012, p. 30)

Sound human resource and budget planning is fundamental to an organization's capability and capacity to effectively carry out evaluations and to deliver and produce reliable information. This category is an essential element of the evaluation policy and is likely to be favourable to ECB because indicates a high level of commitment to the evaluation and to the maximization of the productivity of the organization through the optimization of the effectiveness of its resources.

Evaluation Norms and Standards

This component of evaluation policy emerged from the data as a new category that is not included in the framework for evaluation policy set out in Trochim (2009). I observed that 33 evaluation policies included such elements as (1) evaluation standards, (2) evaluation criteria, (3) evaluation norms, and (4) evaluation principles that must be considered not only when carrying out the evaluation but also when thinking about planning, designing and using the evaluation. Therefore, the data show that this category of the evaluation policy is critical to the creation of a supportive organizational evaluation culture that positively influences the individual and organizational performance of the evaluation as well as the degree of learning that flows from the same.

Organizational culture consists of an organization's vision, norms, values, systems, attitudes, and written and unwritten rules that have been developed over time, are considered valid, and are communicated and reinforced through various means, thus ultimately shaping organization members' perceptions, behaviours and understandings (Needle, 2004). Explicit evaluation policies that include clear norms and principles of evaluation help to foster an organization culture in which members are aware of and agree upon the value of the evaluation function and its role within the organization. Such a culture can improve evaluation practice by guiding and coordinating organization members' conduct towards a vision and towards specific performance goals that benefit the organization. Here is an example:

The following five principles/norms/standards form part of the Institute's evaluation policy framework: **a) Utility:** Evaluation should be planned and conducted with a view to serve the information needs of its intended users; **b) Accuracy and credibility:** Evaluation should be conducted with the necessary professional expertise and be based on the principle of impartiality; **c) Feasibility:** Evaluation should be as practical, politically viable and cost effective as possible, and should take into consideration time and financial and human resource requirements; **d) Consultation, access to information and transparency:** Evaluation should be conducted in a transparent manner with stakeholder consultation and access to relevant information; **e) Propriety:** Evaluation should be undertaken in a legal and ethical manner with regard to the rights and welfare of those involved in and affected by assessments. (UNITAR, 2012, p. 6)

These five norms are present in most of the evaluation policies considered in this study, along with the following set of criteria for evaluation recommended by the OECD Development Assistance Committee: (1) relevance: the degree to which the evaluation

responds to the needs and priorities of the targeted beneficiaries, to the contextual situation or situations to be addressed, and to the priorities of the donor or donors; (2) effectiveness: the extent to which the evaluation has achieved its objectives; (3) efficiency: the cost-effectiveness of transforming inputs into outputs while taking into consideration alternative approaches; (4) impact: the cumulative effect and/or undertakings that may produce positive or negative, intended or unintended changes; and (5) sustainability: the likelihood that the benefits derived from an evaluation will continue over time following its completion.

Some evaluation policies include respective sets of evaluation criteria and evaluation standards and distinguish between the role of each set and the ways in which each influences evaluation practice. Here is an illustration:

Evaluation criteria differ from the evaluation standards and process in that the criteria inform what we evaluate, (the focus of inquiry), whereas the standards and process guide how we conduct the evaluation. The evaluation criteria are complementary, and together they seek to provide a comprehensive evaluation of IFRC's work. (IFRC, 2011, p. 4)

It is clear that this component of the evaluation policy can be used to identify the evaluation norms and standards that are accepted as correct within an organization. Organizations include this component in their policies in order to define the types of qualities that should be considered in the evaluation process. It is important to include norms and standards within evaluation policies because they outline professional expectations, ensure consistency throughout the organization's evaluation process, and provide evaluators and managers with key information about the quality of the evaluation that they must provide. Thus this category is likely to be favourable to ECB because it ultimately establishes measures that allow for effective quality assurance of evaluation performance.

Evaluation Planning and Processes

This category is included in Trochim's (2009) framework for evaluation policy under the designation "evaluation *processes and methods*." Following the review and analysis of the 52 evaluation policies included in this study, I modified this category and renamed it "evaluation *planning and processes*" in order to more closely reflect the study findings. This category comprises evaluation design and planning; evaluation methods; evaluation participation; and evaluation with partners. The modification of the category was based on the finding that planning is a wide-ranging and critical aspect of evaluation and on the

observation that the majority of the evaluation policies include information about planning the evaluation and about the processes of doing the evaluation. This means that planning in evaluation policies is perceived as a crucial and separate step that precedes the conduct of the evaluation.

The data reveal that evaluation planning involves identifying and documenting the evaluation's scope, design, objectives, questions, methods, requirements, deliverables, and resources. It also includes setting a schedule for the implementation of the evaluation. Therefore, evaluation planning is important not only for managing the evaluation time, cost, processes, and quality, but also for ensuring that the evaluations that are conducted produce results that are potentially useful and will satisfy the accountability requirements. Here are two examples:

Evaluations plans need to fulfill the following criteria: (a) Relevant and prioritized to focus on the most important information needs of decision-makers; (b) Oriented towards the relevant results frameworks, in particular to provide strategic coverage of key results; (c) Explicit regarding coordination of evaluation efforts across offices, between agencies and with national stakeholders; (d) Realistic, achievable and monitorable; (e) Timely in terms of scheduling activities, to meet given end uses; (f) Costed and supported by a budget for implementation. (UNICEF, 2013, p. 12)

A prerequisite for evaluation is a coherent results framework. Results frameworks are developed when interventions are planned; they include statements of the expected results and the logical sequence of those results. (UNFPA, 2013, p. 5)

Many policies focus on two dimensions when planning the evaluation: utility (who will use the results and what information will be most useful to them) and feasibility (how much time and resources will be needed for the evaluation).

The evaluation plan shall provide explanations about the selection of evaluation topics yielding benefits for learning and accountability and with high relevance on WIPO direct beneficiaries. (WIPO, 2016, p. 5)

A number of evaluation policies also provide information on the selection of the most appropriate evaluation method. All of these policies indicate that it is the evaluation question and the other key considerations, such as cost, feasibility, and relevance that determine the selection of the evaluation method, rather than the inverse. All of the policies affirm that evaluations must employ data collection and analysis procedures that provide useful and valid information on the effects of a project and that there are multiple evaluation methods and designs that can be employed to accomplish these tasks. For example,

Evaluations will use methods that generate the highest quality and most credible evidence that corresponds to the questions being asked, taking into consideration time, budget, and other practical considerations. . . . No single evaluation design or approach will be privileged over others; rather, the selection of method or methods for a particular evaluation should principally consider the appropriateness of the evaluation design for answering the evaluation questions as well as balance cost, feasibility, and the level of rigor needed to inform specific decisions. (USAID, 2016, p. 8)

The most appropriate design, method, approach and/or technique should be selected, taking into consideration the evaluation question(s), scope, criteria, human and financial resource requirements and availability, as well as guiding principles and good practice standards. (UNITAR, 2012, p. 12)

A number of evaluation policies also outline the expectations for stakeholders' participation in the evaluation process while indicating that these expectations should be considered when planning, designing, and implementing the evaluation. Many organizations' policies recognize that the engagement of stakeholders in designing the evaluation entails strategically thinking through multiple potential methods of data collection and analysis, as well as multiple potential ways in which those methods can be matched to the aims and questions of the evaluation, to the available data, and to the feasibility of implementing the chosen methods with the required degree of rigour. Indeed,

processes should be in place to secure the participation of individuals or parties who may be affected by the evaluation, are able to influence the implementation of recommendations or who would be affected in the long term. Stakeholders should be consulted in the planning, design, conduct and follow-up of evaluations. (UNEG, 2016, p. 24)

The specific primary stakeholders to involve in an evaluation can be determined by their involvement in design and delivery of the program, strategy or project being evaluated. The primary stakeholders should be specified in the Term of Reference, and it is the responsibility of the evaluation manager to ensure that consultations with stakeholders take place. (ILO, 2012, p. 26)

Furthermore, a considerable number of evaluation policies, especially the revised and/or recently established ones, acknowledge that stakeholders' participation in all stages of the evaluation is critical to the creation of a common understanding of the purpose and use of the evaluation and to the development of evaluation capacity within the organization; this is especially the case because stakeholders' participation and the optimization of their level of awareness and understanding of evaluation are essential for fostering a climate of learning within the organization. Two examples are as follows:

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)

Through participation, ILO constituents gain hands-on experience with evaluation and improve their know-how regarding its use. Developing the evaluation capacity of constituents is one of the priorities of the ILO's Evaluation Strategy and training material has been developed for this purpose. (ILO, 2012, p. 26)

Some evaluation policies extend beyond the stakeholders' involvement to include information on the conduct of an evaluation in collaboration with partners. Many organizations use their evaluation policy to communicate their views on working in partnership with other organizations on joint evaluations. The data reveal that 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 within the community; and to increase accountability to the community, funding agencies and stakeholders. For example:

By its nature, evaluation brings a wide range of stakeholders together around common endeavours. 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)

As presented in this section about evaluation policy, the elements of evaluation planning, evaluation design, stakeholder engagement, and partnership in evaluation are all critical for ensuring relevance, ownership, credibility, and usefulness. Most of the evaluation policies that are included in this study demonstrate the importance of these elements by explicitly outlining them in the evaluation policy document. Thus, this category is likely to be favourable to ECB because it enhances the prospects for the usefulness of the evaluation.

Evaluation Utilization

This category of evaluation policy is included in the framework set out in Trochim (2009) and was made evident through the analysis of the data. It focuses on the reporting, dissemination and use of evaluation results. A total of 36 of 52 evaluation policies include a category on evaluation use. In most of these policies, the evaluation utilization category is stated under one of various possible titles: (1) use of evaluation; (2) utilizing evaluation

findings; (3) evaluation usefulness; (4) follow up and use of the evaluations; and (5) using evaluation knowledge.

Many of the evaluation policies included in the sample emphasize the importance of identifying the intended users of the evaluation and engaging them in the evaluation process to increase the likelihood that the evaluation will be conducted in ways that will be appropriate and that will lead to it actually being used. The identification of the intended users of an evaluation is also critical for the management and accountability of the organization as it is considered to be the responsibility of the intended user to use the evaluation findings. Two illustrations are as follows:

Each evaluation shall have an identified owner, such as the responsible officer of a cluster, program, office or project. It is the responsibility of the owner to utilize the findings of the evaluation and develop an action plan and timeline for the implementation of the recommendations. (WHO, 2012, p. 9)

At IDRC, 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)

Several evaluation policies also demonstrate that evaluation use starts during the initial planning stage of the evaluation by involving stakeholders in the identification of priority questions, preferred methods and an appropriate dissemination strategy in order to ensure that the evaluation is as focused as possible on the aspects that will inform and influence stakeholders' future actions and decisions and minimize the risk of generating results that may never be used. Indeed,

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. They should continue to be involved during the course of the evaluation to discuss mid-term findings, when applicable, as well as the actual findings and the draft report. (ILO, 2012, p. 45)

The majority of the evaluation policies also discuss within this policy category the reporting of evaluation findings and indicate that evaluation reports must meet the reporting requirements of the organization and the needs of the primary intended users. The policies recognize that different groups of intended users have varying needs for the evaluation report (i.e., not all evaluation users will need a full technical report); therefore, evaluation plans

should include a communication strategy to effectively 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. (Dutch Ministry of Foreign Affairs, 2009, p. 6)

I 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 with a view to fulfilling the goals of the evaluation. A number of policies outline the role of organization managers in following up on the evaluation and indicate that they need to give careful consideration to the evaluation findings, recommendations, and lessons learned. Policies also provide guidelines for the management's responsibilities as well as descriptions of the key steps to follow up on the evaluation report recommendations. These guidelines and descriptions can be seen in the following illustrations:

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...The Evaluation Department undertakes to monitor the implementation of the follow-up activities at regular intervals. (Danish Development Cooperation, 2016, p. 38)

UNICEF offices must prepare a formal management response and make appropriate arrangements to maximize the use of evaluation results. For each evaluation, the office should: (a) Prepare and implement a multi-method communication plan to disseminate the report and promote appropriate application of the results; (b) Prepare a formal management response, with the active participation of concerned stakeholders, which reacts to the evaluation recommendations; and takes follow-up action to implement agreed recommendations. Where recommendations are rejected or only partially accepted, the rationale should be noted. Regional offices should maintain oversight of the management response process; and (c) Pause program implementation, where necessary, in cases where evaluation results indicate that a proposed program requires modification. (UNICEF, 2013, p. 3)

These examples confirm that communication and accountability are critical for effectively managing the evaluation function and enhancing evaluation utilization. The ultimate purpose of evaluation is to ensure that sound decisions are made; thus, evaluation policy is a vehicle for communicating the expectations, roles and responsibilities of organization members with respect to the use of evaluation for decision-making, for internalizing knowledge sharing to

achieve transparency in the methods used, and for holding employees accountable for their performance.

Some policies also include guidelines and strategies for the dissemination of evaluation findings and documents, specifically as an aspect of evaluation utilization. The main purpose of this component is to make the evaluation results accessible to a wide range of relevant stakeholders, as this will contribute not only to greater accountability for the organization but also to broader knowledge generation and mobilization. It is especially important for development organizations to have accountability requirements and transparency commitments to make evaluation products accessible and barrier-free to each of the major groups of users. 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; each management response is uploaded in the tracking system within 30 days of final approval of the management response document; followed by quarterly updates on implementation of agreed actions. (UNICEF, 2013, p. 14)

These examples demonstrate that the category of evaluation utilization is meant to provide clear guidelines that underscore the importance of the various types of evaluation use (including conceptual use and process use) and that ultimately lead to the instrumental use of findings in improving decision-making and improving programs. The encouragement of conceptual use is noted in policies that highlight the importance of evaluation users' involvement in increasing awareness and understanding of program issues and addressing them; this in turn increases the likelihood that informed decisions will be made and that beneficial changes will occur.

An emphasis on process use is specifically noted in policies that outline the benefits and consequences of stakeholders' involvement in the evaluation process in terms of increased learning and increased ownership. At the same time, an emphasis on instrumental use is noted in policies that provide details on the implementation of evaluation findings and the application of recommendations. I observed that the majority of policies indicate that the identification of evaluation use starts with the planning stage of the evaluation and requires that every evaluation plan identify users, involve them in all of the evaluation processes, and indicate how the results are to be communicated, used and disseminated. Therefore, this

category is likely to be favourable to ECB because it outlines the expectations, roles, and responsibilities of organization members with respect to the use of evaluation for making decisions and internalizing knowledge sharing.

Meta-Evaluation

This category of evaluation policy focuses on the review and evaluation of evaluations (and in some cases evaluators) as part of the organization's evaluation management and quality control system. The meta-evaluation category is included in Trochim's (2009) framework for evaluation policy and is identified in 23 policies of various types of organizations under such sections as: (1) quality and monitoring of evaluations; (2) quality assurance; (3) quality control; and (4) quality standards. Moreover,

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)

Meta-evaluation policy relates to the quality control of evaluation processes and 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.

Meta-evaluations of the evaluation process should be periodically conducted to: take inventory of evaluations and synthesize evaluation results; examine evaluation methodologies; check compliance with and consider revision of the evaluation framework; inform corporate policy and strategy in the selection of evaluation exercises; and improve dissemination and utilization of lessons learned from evaluations. The planning and evaluation department will lead this practice at a minimum of every two years. (French Development Agency, 2011, p. 6)

The data reveal that meta-evaluation is widely used in international development organizations and is independently undertaken based on a set of critically recognized evaluation standards expressly developed for this purpose, as is illustrated in the following examples:

How can an evaluation be judged that it was carried out professionally? How can an evaluation be assessed formally? The Development Assistance Committee of the OECD has developed quality standards for that purpose. (Austrian Development Agency, 2009, p. 31)

Independent quality assurance of evaluation at design, inception and draft final report stages is mandatory for all evaluations where DFID is the main funder. Independent quality assurance is essential to ensure evaluations meet quality standards throughout the evaluation life cycle from inception and design to final report and communication. (DFID, 2013, p. 15)

Quality control is exercised throughout the evaluation process. The quality of an evaluation report is assessed and rated against a set of criteria. (UNIDO, 2015, p. 11)

It should be noted that meta-evaluation is considered one of the many aspects of the quality assurance system in some organizations. Such a system is used as a formal management system and is intended to raise standards of work and increase an organization's credibility by demonstrating its accountability to various stakeholders. This can be seen in the following example:

Key elements of the quality assurance mechanisms include: (a) guidance and tools for conducting and managing evaluations; (b) approval by the Evaluation Office of all terms of reference for evaluations; (c) centralized pre-qualification of all evaluators; (d) centralized, high quality criteria for assessing all evaluation reports; and (e) a quality assessment of all evaluations, processes and the evaluation function commissioned by the Evaluation Office. (UNFPA, 2013, p. 13)

The inclusion of a category on meta-evaluation within the evaluation policy indicates that the organization is committed to a quality assurance system and is striving for continuous improvement of the way it manages and delivers its services and other activities. It also promotes the utilization of evaluation by involving the primary intended users and other key stakeholders in a review of the evaluation. The following statements illustrate this commitment:

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. (Spanish Agency for International Development Cooperation, 2013, p. 16)

At the end of each biennium the Evaluation Office will prepare a Biennial Evaluation Report. This report will summarize the performance of the organization through trends and patterns observed during the biennium from completed evaluations at all levels. The patterns and trends will be used to identify recommendations and lessons to be brought to the attention of, and discussed with, UNEP Senior Management. (UNEP, 2009, p. 8)

The meta-evaluation category is found to be critical to the creation of a supportive organizational setting that influences organizational learning and performance management. It is clear that this category is likely to be favourable for enhancing ECB.

Evaluation Capacity Building

This category focuses on the range of efforts that an organization may undertake in order to integrate evaluation practices and use them at all levels of the organization and thus strengthen its operations and improve its activities. The data reveal that many organizations acknowledge the importance of strengthening the evaluation capacity of individuals, organizations, communities, and national governments as a key to managing evaluation and improving transparency and mutual accountability. The ECB category is included in Trochim's (2009) framework for evaluation policy and is explicitly identified in 17 policies included in the sample for this study. It should be noted that only nine of the policies have a specific section for ECB; in general, ECB-related policies are sprinkled throughout the policy documents and are presented under different sections such as: (1) evaluation goal; (2) evaluation management; (3) quality assurance; (4) roles and responsibilities; and (5) evaluation utilization.

While some policies point to capacity building undertaken by an organization or individual with particular expertise in transferring knowledge and skills to others—such as, for example, “UNIDO enhances its organizational capacity and the competence of its staff by fostering the exchange of experience and identifying good evaluation practices at the international level” (UNIDO, 2015, p. 11)—in the data there are many specific examples where the ECB activities are not just about developing skills for doing evaluation but also about effectively managing, performing and using evaluations as well as developing an evaluation culture of valuing evidence and evaluative thinking. For example:

An organization may support capacity development by improving evaluation knowledge and skills, strengthening evaluation management, stimulating demand for and use of evaluation findings, and supporting an environment of accountability and learning. (OECD, 2010, p. 7)

UNESCO remains committed to building its own internal capacities in evaluation. The Evaluation Office provides programme staff with relevant training and guidance material to strengthen the necessary skills and knowledge required to carry out self-assessments and to manage decentralized evaluations. The conduct of formative evaluations whereby UNESCO staff participates more closely in all phases of the evaluation is another method for building

an evaluation culture and evaluation capacities. The Office is also working to institutionalize a system of monitoring and evaluation focal points across the Organization, along with a customized training programme, aimed at strengthening internal capacities and improving the quality of evaluation across the UNESCO system. (UNESCO, 2015, p. 16)

UNHCR recognizes the need to develop an organizational culture that is amenable to the development of an effective evaluation function. In that respect, the Office is striving to ensure that managers and staff members are fully committed to evaluation and learning, and that using evaluation to bring about change is positively encouraged throughout the organization. (UNHCR, 2010, p. 1)

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)

Moreover, the data reveal that many of the development agencies considered in the sample are focusing on organizational and individual capacities for evaluation in partner countries, with aims that include strengthening domestic accountability, facilitating collaboration with partners, and improving development effectiveness at the country level. For instance,

there is growing interest in strengthening the evaluation capacity of national partners. The revised Evaluation Policy seeks to address these issues and provide a framework for future improvements.

Developing the evaluation capacity of constituents is one of the priorities of the ILO's Evaluation Strategy and training material has been developed for this purpose. (ILO, 2012, p. 2)

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 (New Zealand Aid Program, 2014, p. 5)

While this is sometimes achieved through targeted training or other specific capacity-building programs, a number of 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. For example:

UNFPA evaluations are carried out in a manner that enhances national capacity through the participation of governments and key stakeholders in evaluations; through support for country-led evaluations; and through the use of national evaluation systems. Capacity-building initiatives include guidance, training, and the sharing of good practices and lessons learned. . . . Partnerships with evaluation networks, professional bodies, national and regional institutions, United Nations partners and other development partners will further reinforce evaluation capacity. (UNFPA, 2013, p. 13)

Enhancing the capacity of evaluation among our partners is achieved through championing the central role that evaluation capacity development plays in DFID's development programmes to strengthen the enabling environment for evidence-based policy. This supports partner agencies, Government, Non-Government and private actors to maintain or build strong evaluation systems, and enhances supplier capabilities to respond to the demand. (DFID, 2013, p. 14)

Evaluation contributes to social transformation, not only through the use of its products, but also as a catalyzing process for changes at the personal, organisational and systemic levels. To favour this transformative potential, Spanish Cooperation will support capacity building for the different actors involved, promote their effective participation in the different phases of evaluation processes, and promote dialogue, critical reflection and the review of practices and assumptions (Spanish Agency for International Development Cooperation, 2013, p. 4)

Moreover, the data show that the primary goals of development agencies in strengthening evaluation capacity in partner countries are to inform policy making, to benefit from effective partner participation in the evaluation process, and to cultivate a receptive environment for implementing recommendations and enhanced accountability mechanisms to manage development results. Indeed,

capacity building linked to the evaluation process will guide relations among the actors of the Spanish cooperation system and between Spanish Cooperation and its partners. The use of the evaluation systems of partner countries will be enhanced and support will be provided for the development of local capacities, both at the institutional level and through evaluation associations and networks. (Spanish Agency for International Development Cooperation, 2013, p. 7)

Some policies indicate their commitment to the ECB process by specifying the budget for its activities. For example, at UNESCO,

the target for funding evaluation activities related to extra budgetary projects is set at 3% of project budgets. Two-thirds of this allocation is to be used for decentralized evaluations, project evaluation capacity-building, monitoring and evaluation, national evaluation capacity building, while one-third is to provide funding towards cross-cutting corporate evaluations, dissemination of lessons learned and synthesis work. (UNESCO, 2015, p. 17)

Furthermore, some policies explicitly assign ECB roles and responsibilities to members and units of the organization:

Directors implement and support actions to strengthen evaluation capacity and engagement in the headquarters units they oversee and the global networks they manage. (UNICIF, 2013, p. 9)

Strengthening OCHA's internal evaluation capacity means building evaluative capacity throughout the organization. It also requires a cultural change in attitudes, which values the contribution of evaluation to OCHA's broader management reform framework and agenda. In

this context, the Evaluation Section will: i) organize briefings and orientations on specific issues related to OCHA's overall monitoring and evaluation framework and will also prepare detailed operational guidelines for the conduct of evaluation activities by OCHA; ii) work closely with the Staff Development and Learning Section to integrate lessons learned from all evaluations as well as relevant M&E concerns into all staff training programs and iii) to the extent that it has the capacity provide a Help Desk function to other units, sections and branches of OCHA undertaking their own reviews and assessments. (OCHA, 2012, p. 4)

Thus, the findings of the analysis indicate that, despite being included in only about one third of the study sample, ECB is an important component of evaluation policy that aims to improve transparency and build a performance culture within organizations (and national governments), to support results-based management and policymaking, and to strengthen accountability relationships—by providing support for the creation or strengthening of organizational/national monitoring and evaluation systems.

Competencies and Ethics

This is a new category of evaluation policy that emerged from the data and is not included in the framework for evaluation policy set out in Trochim (2009). It focuses on the qualifications, skills, expertise, and ethics that evaluators require in order to fulfil their roles and responsibilities within the evaluation process and thus ensure the credibility and quality of the process. I observed that 15 evaluation policies included such elements as (1) evaluator qualifications, (2) evaluation competencies, and (3) evaluation professionalism and ethics—elements that must be considered when recruiting evaluators as well as when conducting, reporting and using an evaluation. It can be argued that success in evaluation hinges on recruiting a competent evaluator, performing the evaluation with due regard for ethics and with due diligence, and ensuring that the evaluation results are accurate and useful and that the decision-making consistently reflects the organization's values and ethical principles. The data show that there is an overlap between competencies and ethical principles when it comes to practising evaluation. While ethics constitute the fount from which the values for all decision-making flow, evaluators with the right skills and competencies play a critical role in implementing and enforcing ethics in the process of generating the information required for sound decision-making. Therefore, it is evident through the analysis that addressing evaluators' competencies and ethics is an important component of evaluation policies, especially given the increasing complexity of programs and policy contexts. Indeed,

all those engaged in designing, conducting and managing evaluation activities should aspire to conduct high-quality work guided by professional standards and ethical and moral principles. This includes heads of evaluation offices/units, evaluation office staff, decentralized evaluation staff, evaluation managers and evaluators. (UNEG, 2016, p. 19)

High-quality evaluation requires sound technical and management skills. This requires that staff, consultants and other personnel engaged in designing, conducting and managing evaluation activities possess core evaluation competencies; evaluation specialists, whether staff or consultants, have the relevant educational background, qualification and training in evaluation, as well as relevant experience; evaluation staff have access to appropriate training and support; and project staff and managers receive the training and technical support required to fulfill their evaluation accountabilities. (UNICEF, 2013, p. 16)

This section identifies what CIDA is looking for to ensure satisfactory performance of contract in terms of the evaluation team's qualifications, experience, expertise and demonstrated abilities. Such requirements may include: experience in conducting evaluations, with a focus on previous assessments of development cooperation programming and/or work performed in the subject region or country; language requirements; technical and analytical expertise; the involvement of local firms and/or individuals; and experience in conducting participatory evaluations. (CIDA, 2013, p. 10)

The correct identification of the competencies required for evaluation-related positions is a crucial aspect of the development and improvement of the evaluation system within the organization. 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 and ethics. For instance:

This policy requires that Evaluation Office staff and evaluation consultants must have personal and professional integrity; respect the right of institutions and individuals to provide information in confidence; ensure that sensitive information cannot be traced to its source; take care that those involved in evaluations have a chance to examine the statements attributed to them; be sensitive to beliefs, manners and customs of the social and cultural environments in which they work; not have been involved in the process of development, implementation or supervision of the program projects or policies being evaluated to avoid conflicts of interest; be sensitive to and address issues of discrimination and gender inequality; discreetly report evidence of wrongdoing to the appropriate investigative body. (UNEP, 2009, p. 5)

The Director of the Evaluation Group must have proven competencies in the management of an evaluation function. 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)

I noticed that some 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. Here is an example:

The delivery of DFID's Evaluation Strategy depends on having a strong cadre that meets the needs of the decentralized evaluation program and includes staff who can provide leadership across all areas of the evaluation life cycle. The evaluation accreditation model and evaluation competencies are designed to ensure high professional standards are maintained, and to ensure staff skills are aligned with the requirements of this strategy to strengthen the quality of DFID evaluations. (DFID, 2013, p. 11)

Moreover, some policies provide more details about specific evaluator competencies. For example,

the qualifications and skill areas to be specified could include areas of technical competence (sector, issue areas); language proficiency; in-country or regional work experience; evaluation methods and data-collection skills; analytical skills and frameworks, such as gender analysis; process management skills, such as facilitation skills; gender mix (not to be confused with gender analysis skills). (ECLAC, 2014, p. 13)

I also observed that some policies refer to the principles or standards of other organizations and/or professional networks and indicate their commitment and adherence to them. This can be seen in the following examples:

Evaluations must be conducted in an ethical and legal manner, with particular regard for the welfare of those involved in and affected by the evaluation. Evaluations should abide by professional ethics, standards and regulations to minimize risks, harms and burdens to evaluation participants. . . . Evaluators should respect the customs, culture, and dignity of human subjects. This includes differences due to religion, gender, disability, age, sexual orientation and ethnicity. Particular attention should be given to address issues of discrimination and gender inequality, (in accordance with the United Nations Universal Declaration of Human Rights). (IFRC, 2011, p. 5)

DFID will uphold the UK obligations and commitments to gender equality and human rights, making sure that these are increasingly mainstreamed into all of its evaluation work. (DFID)

The Canadian Evaluation Society's "Guidelines for Ethical Conduct" address the importance of: 1) competency in the provision of service, 2) integrity in relationships with stakeholders, and 3) accountability for performance and product. (CIDA, 2013, p. 10)

The analysis reveals 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)

The universally recognized values and principles of human rights and gender equality need to be integrated into all stages of an evaluation. It is the responsibility of evaluators and evaluation managers to ensure that these values are respected, addressed and promoted, underpinning the commitment to the principle of “no-one left behind.” (UNEG, 2016, p. 12)

Unless ethical standards are respected, the credibility and quality of an evaluation may be placed in jeopardy. Thus, by clearly communicating their ethical principles within and through the evaluation policy, organizations leave little room for doubt, confusion, or contradiction. Therefore, this category of competencies is likely to be favourable 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.

Other Findings

The analysis of the data revealed that there is considerable variation among the evaluation policies included in the sample in terms of their respective depths, lengths, and components. This variation can be attributed to each organization’s type, size, and scale of operations. Although the 10 evaluation policy categories identified and discussed above represent a comprehensive framework and capture almost all of the components associated with critical aspects of the evaluation process, some policies include unique components that are worth mentioning and documenting in this section. These unique components are set out hereunder.

(1) Purpose of the evaluation policy: A total of 11 organizations started with a clear statement of the purpose of the evaluation policy. I observed that such a statement is important for introducing the reader to the goals and significance of the document. Some examples:

The objectives of this policy are to: (i) improve the achievement of results across government; (ii) enhance the understanding of the results government seeks to achieve, does achieve, and the resources used to achieve them. (Policy on Results, the Government of Canada, 2016, p. 1)

The purpose of this policy is to define the overall framework for evaluation at WHO, to foster the culture and use of evaluation across the Organization, and to facilitate conformity of evaluation at WHO with best practices and with the norms and standards for evaluation of the United Nations Evaluation Group. (WHO, 2012, p. 4)

The overall aim of this policy is to establish common structures and standards across the IUCN Secretariat that govern the application of effective monitoring and evaluation systems with a view to maximizing the benefits from IUCN interventions. More specifically, this policy aims to: Demonstrate IUCN's commitment to monitoring and evaluating its work and using the results to drive performance and impact; Set out minimum requirements, principles to be respected, as well as roles and responsibilities. All IUCN staff must comply with this policy. (IUCN, 2015, p. 2)

Thus, a statement of the purpose of the policy can be included in order to describe its objectives and how it will inform evaluation capacity building in terms of conducting evaluation and using it.

(2) Evaluation history: A total of eight policies included information about the history of the evaluation function within the organization. I observed that organizations include this kind of information in the policy in order to indicate changes in the system, explain the reason or reasons for developing/creating the policy, or discuss lessons learned from previous experience. Here is an example:

In 2002, WHO developed a framework on Programme Management, focusing on monitoring, evaluation and reporting. This guidance outlined evaluation in the context of the principles of results based management...However, from a broader institutional perspective, it has been less successful in fostering an evaluation culture, developing evaluation capacity at all levels of the Organization, and in promulgating participatory approaches to evaluations. The causes for this include the institutional arrangements for the evaluation function (including a lack of a direct mechanism for oversight by the governing bodies) and the absence of an effective budgetary allocation mechanism to resource the evaluation function. (WHO, 2012, p. 1)

This aspect is helpful for documenting improvements and other changes to the evaluation function to help the organization members understand directions in which the policy has been improved and added value.

(3) Importance of evaluation: Many organizations included a statement about the importance of evaluation and how it influences the organization. This is especially helpful for communicating to stakeholders how evaluation is perceived within the organization. For example,

Evaluation is at the heart of the General Secretariat for International Development Cooperation's efforts to more effectively and more efficiently meet the ultimate purpose of Spanish Cooperation: contributing to human development, reducing poverty, and achieving the full exercise of rights. (Spanish Agency for International Development Cooperation, 2013, p. 1)

Evaluation is a key part of the evidence agenda and is one of several related evaluative activities undertaken across the organisation... Evaluation complements other functions that

aim to achieve the goal of strengthening the evidence underpinning DFID policies and programmes to ensure development results and impact along with ensuring value for money of DFID spend. (DFID, 2013, p. 2)

Evaluation looks at whether the right things are being done and whether they are done correctly. It examines the rationale and the justification of activities. It assesses the effectiveness of achieving expected results, and it examines the efficiency of the use of inputs to achieve results. Finally, evaluation recommends and suggests better ways of achieving expected results. (New Zealand Aid Program, 2014, p. 2)

(4) Related policies: A number of evaluation policies include sections that mention other policies that affect or are affected by the evaluation policy. This is important for ensuring that there are no conflicts between different policies and procedures when implementing the evaluation policy within the organization. For instance,

This update had been made to ensure consistency with revisions to USAID's Automated Directives System (ADS) Chapter 201 Program Cycle Operational Policy, which was released September 2016. The ADS revisions changed evaluation requirements to simplify implementation and increase the breadth of evaluation coverage. (USAID, 2016, p. 1)

(5) Examples/templates: Some evaluation policies provide examples and templates of various parts of evaluation reports in order to show what they are looking for in future evaluation reports in terms of components and contents. SIDA, for example, offers the following guidance:

This format is intended to help guide the structure and main contents of evaluation reports commissioned by Sida. It is not compulsory, but should be used if there is no particular reason for doing otherwise. (SIDA, 2014, p. 84)

Some policies also provide examples of various data collection methods and their respective requirements as these can help in planning and designing the evaluation.

(6) Rationale for the update: A total of 11 out of the 17 organizations that revised their policies included a statement about why they did so. This information is helpful for clarifying the reason or reasons for the update, reflecting improvements and indicating changes in the system. For instance,

UNICEF has revised the Evaluation Policy in order to respond to changes within the organization and its operating environment. In recent years, significant changes have been taking place in the broad field of development cooperation and in the work of the United Nations. Expectations are rising regarding the transparency, accountability and effectiveness of development processes and the efficient use of resources allocated for development. Evaluation methods and approaches have also been enhanced to address emerging development needs and demands. (UNICEF, 2013, p. 1)

In general, an evaluation policy is the organization's statement of intent with respect to evaluation. It is a communication tool that points out processes, standards, requirements and expectations to various audiences, as can be seen in the case of the IFRC:

This framework is intended for two audiences. First, it is intended to guide people involved in commissioning, managing, or conducting a Secretariat evaluation. This includes those responsible for programme and project management, policy development and review, strategic planning, evaluation capacity building and training, and the evaluators themselves. Second, this framework is intended to inform and assure other stakeholders in the evaluation process, i.e. donors and beneficiaries, of key practices and ethical commitments expected from IFRC evaluation work. (IFRC, 2011, p. 1)

Thus, the elements included in the policy not only reflect the organization's mission, vision, and values, but also shape and establish its organizational culture. An organization that uses evaluation policy to encourage continuous capacity building development recognizes that these categories influence each other and work together to ensure that learning and evaluation activities feed into the organization's goal and allow it to transform continuously for the better.

A New Definition of Evaluation Policy

The review and analysis of 52 evaluation policies developed by various types of organizations provided me with a broad and comprehensive perspective on the ways in which evaluation policy is conceptualized within an organization system. It also allowed me to identify the main and most important categories that organizations include in their policies in order to guide and manage their evaluation practices. These categories cover all of the stages of evaluation, from thinking about the evaluation to proceeding with planning it, designing it, conducting it, managing it, collecting data for it, analyzing it, reporting it and utilizing it as well as disseminating its findings to various stakeholders. The policies also include norms and standards that are critical to establishing the organizational structure and culture that are necessary for supporting evaluation integration within the system of the organization.

The findings of this study led me to think about the current one-and-only evaluation policy definition that one can find in the evaluation literature—the definition that was provided by Trochim (2009) and is included in the first chapter of this thesis. This definition is as follows: “Evaluation policy is any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation” (Trochim, 2009, p. 16). Based on the

findings of this study and, in particular, on a closer examination of the elements of evaluation policy that this definition captures, I decided to reflect on the definition and refine it to make it more inclusive and comprehensive. This study highlights the fact 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 this 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, a revised definition of evaluation policy that captures its actual and practical meaning within an organization is as follows:

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.

I believe that this definition of evaluation policy is more inclusive and comprehensive and is a more accurate reflection of the underlying meaning of evaluation policy and its role in setting the evaluation processes and practices that encourage individuals—and the organization as a whole—to improve their capacity and increase knowledge, competence, and performance.

Summary

In this chapter I provide a comprehensive review and analysis of an extensive sample of evaluation policies as a way to understand current trends and to identify key categories in evaluation policy in relation to organizational capacity to do and use evaluation, and to answer the first overarching question of my research and inform Study 2. Through my analysis of 52 evaluation policies of various organizations, I was able to (1) validate and improve the existing evaluation policy framework developed by Trochim (2009) and identify 10 broad categories for evaluation policy that I believe capture the dynamic range and multiple manifestations of evaluation processes in various organizations; and (2) provide a comprehensive definition of evaluation policy that reflects its meaning within the organization and improves upon the existing definition found in the evaluation literature.

Chapter 5: Findings from Interviews with Evaluation Scholars and Practitioners

The purpose of this chapter is to set out the findings from the interviews that I conducted with 18 evaluation scholars and practitioners. It should be noted that the interview data touch on each of the three overarching research questions of the thesis on: (1) the role of policy in organizational capacity for evaluation; (2) how policy influences the integration of evaluation into the organizational culture; and (3) the role of contextual factors and barriers.

Based on the analysis of my findings, I developed a three-level ecological framework that illuminates the multiple interconnected variables/conditions that moderate the relationship between evaluation policy and organizational capacity to do and use evaluation. This framework, appearing in Figure 4, comprises the following: the *macro* level (which focuses on the social and political context); the *meso* level (which focuses on the organization level); and the *micro* level (which focuses on evaluation stakeholders). The framework represents a response to the research questions posed and provides an organizing schematic through which to illustrate the multiple variables and conditions that are involved in the relationship between evaluation policy and organizational capacity to do and use evaluation as a way of advancing scholarly and professional understanding of the influence of evaluation policy on evaluation practice.

The development of this three-level ecological framework was informed by Bronfenbrenner's ecological system theory (1979). This theory is well suited to understanding evaluation capacity development because of its focus on providing a systematic approach to explaining social development and its emphasis on the importance of each level for the development process. The theory recognizes the influence of context, community, and culture on the process of development. Rather than ignoring these fundamental variables, the ecological system theory focuses on explaining the interrelationship among multiple variables of the development process and acknowledges their interdependence, a perspective that helps to promote a holistic understanding.

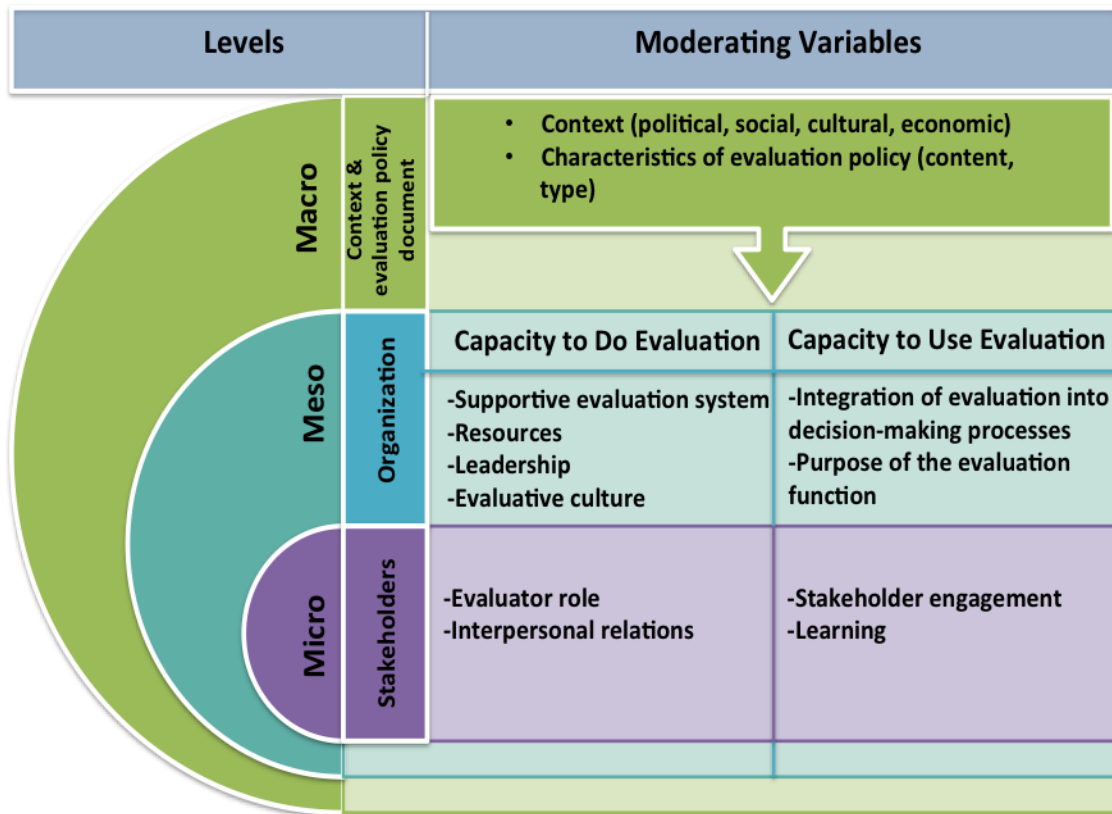


Figure 4. Moderating variables that explain the relationship between evaluation policy and organizational capacity for evaluation.

The theory's flexibility in terms of its inclusion of a varying number of levels as well as an unlimited number of variables highlights its usefulness within the context of this study. I began this analysis by trying to understand the relationship between evaluation policy and ECB within the organizational context. This relationship cannot be understood in isolation, without due consideration of the context; therefore, the ecological system theory provides a useful lens through which to make connections between the different variables involved at different levels of the ECB process.

I use this three-level ecological framework as a means of organizing and presenting my findings throughout this chapter.

Major Findings

The findings reveal that the role of evaluation policy in leveraging organizational evaluation capacity has been quite limited, though this was not always the case. Most interview participants agree that, in and of itself, evaluation policy has little relationship with

an organization's capacity to implement evaluation or to use it. The extent to which evaluation policy positively affects capacity depends on a number of conditions or moderating variables. From my analyses I identified 12 such variables: 1) context (political, social, cultural, economic); 2) characteristics of evaluation policy (content, type); 3) supportive evaluation system; 4) resources; 5) leadership; 6) evaluative culture; 7) evaluator role; 8) interpersonal relations; 9) integration of evaluation into decision-making processes; 10) purpose of evaluation function; 11) stakeholder engagement; and 12) learning.

The conditions or variables that moderate evaluation policy's influences on capacity were found to be operating at one of three levels, as is set out in Figure 4. At the macro level, the social and political contextual conditions and characteristics of often externally imposed policies enhance or impede these policies' influence on organizational capacity to do and use evaluation at the organizational—or meso—level as well as at the stakeholder—or micro—level. The data reveal that the various macro-level conditions have different moderating influences on the capacity to do and the capacity to use evaluation at the organizational level and at the stakeholder level.

As shown in Figure 4, evaluation policy's impact on the capacity to do evaluation at the organizational level is affected by the existence of a supportive system, the availability of resources, leadership commitment, the presence of an evaluative culture that accepts and values evaluation, and the plan and processes for implementing the evaluation policy. These factors have a direct influence on the role of the evaluator and on interpersonal relations among stakeholders as individual-level variables moderating policy effects on the capacity to do evaluation.

On the other hand, evaluation policy's influence on the capacity to use evaluation at the organizational level is influenced by the integration of evaluation into decision-making processes, a key condition that helps to leverage action. The other notable moderating variable is the conceptualization of the purpose of evaluation. At the individual level, stakeholders' engagement in the evaluation process and their learning from evaluation activities are the two conditions that were identified as moderating the influence of policy on their capacity to use evaluation.

In the process of addressing the research questions for the thesis, I considered the macro-meso-micro framework described in this chapter as an interrelated whole. In what

follows, I provide an explication of each level of the framework, including related factors associated with each level and illustrative examples drawn from interviews with evaluation scholars and practitioners. The three-level framework thus provides an organizing framework in which to relate my interviews findings.

Macro Level: Socio-political, Historical, Cultural and Economic Context

In evaluation, context is significant and evaluation capacity is a context-dependent process. The contextual factors affect the development, implementation, and outcomes of policies and programs because they shape the way in which evaluation is approached, practised, and used. Many theorists in the evaluation field have recognized the role played by various contextual factors in shaping evaluation practices. For example, more than 40 years ago Weiss (1973) made the case for the importance of the political context and the ways in which it affects evaluation practice as well as the ways in which evaluators disseminate and communicate evaluation findings in order to foster use. Furthermore, in utilization-focused evaluation, Patton (2008) stresses the importance of evaluators who are embarking on their evaluations conducting situational analysis in order to ensure that they understand the context. The findings of my study indicate that the same applies to evaluation policy.

The macro level focuses on the external factors that affect the role of evaluation policy in shaping evaluation practices and organizational capacity for evaluation. This level highlights the critical influence of two moderating factors: (1) the contextual factors, whether they are political, social, cultural and/or economic, within which the evaluation policy is created and applied; and (2) the characteristics of the evaluation policy itself (i.e., the content of the policy, the type of information). It should be noted that in most cases evaluation policy is externally imposed on organizations by agencies such as governments, public agencies, legal authorities and so on. The organization's external environment influences the role of evaluation policy through the demands for evaluation that it places on the organization, and through the incentives that it provides for conducting evaluation.

The data from the interviews reveal that the **socio-political context** is a critical factor that has a great influence on the role of evaluation policy and how it affects evaluation practice and use. On the one hand, many participants shared the view that externally imposed

evaluation policy creates demand for evaluation and provides the underlying rationale for the role of the evaluation function, as is implied by the following comments:

When evaluation policy is imposed on organizations it can provide a trigger. It provides a basis and provides some legitimacy. In some organizations, it is troubling to get an evaluation going unless you provide some legitimacy, something evaluators can point to... It provides the basic instruction for evaluators and evaluation in an organization, which is probably needed in most organizations. (IP4)¹

An external factor like evaluation policy works better than an internal factor because then you can say, okay we want to keep getting our money from Congress so we've got to make a case that scientifically this stuff is working. (IP11)

One of the participants talked about how evaluation policy in the government context has informed evaluation practice and how the policy requirements have increased the capacity to use evaluation for decision making:

Evaluation policy can outline very specific requirements on how the evaluation should be conducted. I'm here thinking of the federal evaluation policy that identifies the five core issues. Despite the criticism, it tells all the federal government evaluators to ask these five questions. So it is really specific on how to do evaluation and then in the other end of the spectrum it also explains how evaluation is to be used or in what context by specifically stating who are the main intended users of the evaluation. So I think this example shows that evaluation policy, in the federal context at least, covers quite a few elements of the capacity to do and use evaluation. (IP10)

Other participants mentioned that evaluation policy is likely to have a major impact on evaluation practice and capacity building in developing country contexts because the needs and demands for evaluation in these political contexts are more focused on supporting the development of evaluation capacity and on strengthening related systems of management, learning and accountability, as is suggested by the following quotations:

If you want to develop evaluation capacity in a sub-Saharan African country you better have an evaluation policy that tells people there what evaluation is and what they need to know and what they are supposed to do. They need something to refer to; it will guide them through the process. (IP8)

In countries where there is not a real experience or knowledge of evaluations, evaluation policy can really help. It can be a stimulus to get things going. (IP3)

If we talk about Saudi Arabia or Iran or Iraq or to some extent Russia, if the government really wants to get things going, policy or regulation could be a positive thing there. It will function as a kind of new window of opportunities to learn from evaluation. (IP17)

¹ "IP4" is an abbreviation of "Interview Participant #4"; this coding system is used to identify responses of the 18 participants while protecting their identity.

In many African countries now they do talk about evaluation capacity building or developing. They are developing some internal capacity within the government to do evaluation and then try to educate managers about why they should care about evaluation. I've worked in Africa and I know that evaluation policy can help in that regard but you're going to see some differences in a country like Tunisia or Egypt. Tunisia has just developed an evaluation policy that is really going to make a change. (IP5)

On the other hand, many participants shared that in the global North (i.e., developed countries), evaluation policy is mostly perceived as another level of bureaucracy and an additional requirement that actors need to fulfil. This is especially the case in political contexts where there is a high demand for evaluation for accountability purposes, as is implied by the following comments:

In Europe, we have evaluations, we have performance audits, we have oversight, we have supervision, we have inspections and in the 80s there was a big difference between oversight and inspection on the one hand, and evaluation on the other hand. Those were under-developed, now it is considered to be all part of the evaluation machine. Whether you are evaluated by an inspector or by an oversight agent or by a supervising organization, it is considered more or less the same. We are evaluated to death. (IP2)

In England, the government over-evaluates everything. Evaluation became a burden and once you say evaluation people cannot stand you. I know academics in England who left the profession because they are overwhelmed by all of these evaluations and policies and there are so many forms that they need to fill out and they can barely teach. (IP15)

As soon as you say the word "policy" in Israel people will get excited! They will be like what is this policy going to do to me? Policy is a very threatening word so you take policy and put it with evaluation together. It is not appealing to most people. Evaluation is very political there. (IP9)

It seems likely that in contexts where evaluation is mainly used for performance measurement and performance accountability, evaluation policy becomes the dominant accountability tool. This can hinder learning, improvement and, ironically, performance, as is suggested by the following comments:

If policy is not well thought out it can damage practice. You know in England strict policy that focuses mainly on accountability was a disaster because it was driving teachers out of the profession. It was making vast use of time and resources but researchers found it very difficult to link outcomes with cost because it was such a high stakes accountability system. (IP1)

We are living in the Western world in an almost over-evaluated society. Do you know the Swiss economist Bruno Frey? He coined the word "evaluaitis" as a kind of infection that you experience because everything has to be evaluated all the time. In that situation, an evaluation policy is very difficult to contribute in a positive way. (IP6)

Some participants also pointed to the influence of **the historical and economic aspects** of the evaluation policy context and the ways in which they affect the practice of evaluation and people's perceptions of evaluation. For example, two participants from Europe argued that the European Structural Funds were a major driver for spreading the practice of evaluation throughout Europe in the 1980s and indicated that the introduction of evaluation into many European countries was a result of the requirements of the structural fund regulations to assess value for money, requirements that continue to influence how people think about evaluation. In such contexts, evaluation is associated with a high level of accountability, with the main motivation being the fulfilment of external requirements accompanied by little or no interest in organizational improvement or learning. Indeed,

I would argue that evaluation began in Ireland sometime around the late 1980s when we had a very large injection of funding from the European Union called the European Structural Fund and with them there came the demand for evaluation and all of these evaluation policies. Up until then we did not really evaluate things and, even if we did, there was really no evaluation units anywhere. Overnight suddenly everything had to have evaluations so we developed these policies and all government departments and all state agencies began to do evaluations mostly for the value of money to meet the policy requirement. (IP15)

Many people still have negative associations with evaluation here in Europe and now we're 30-40 years later and evaluation still makes people feel distrustful. (IP2)

The other moderating macro-level variable that affects how evaluation policy influences organizational capacity is **the characteristics of the evaluation policy itself** (i.e., content, type). The data reveal that the type of information and the material contained within the evaluation policy in relation to the policy's requirements, whether the policy is descriptive or prescriptive, have a significant impact on evaluation practice and on evaluation capacity. The data show that descriptive evaluation policies are more favourable for guiding evaluation practice because they provide general descriptions and guidance for the evaluation function without restricting practice. Prescriptive policies, on the other hand, include extensive details and instructions that restrict evaluation practice and reduce the flexibility needed to learn from evaluation and improve practice and capacity. As some participants explained,

Policymakers are thinking of themselves as closer to gods, and they think, well, they can specify evaluation policy any way they want. I'm saying that no, there should be some constraints on the policymaker. It's sort of like in medicine, the do-no-harm criteria. Policymakers get in there and muck around too much. The last thing I want them to do is to

micro-manage evaluation because this is when evaluation policy becomes less a shaping mechanism and more onerous and punitive and burdensome. (IP13)

Evaluation policy at the federal government is considered to be very prescriptive and it results in evaluations that are so focused to meet the requirements of the policy that they set aside the actual informational needs of stakeholders. So evaluators really struggle because the policy is so explicit about the fundamental issues and the key evaluation questions to be asked, they are so focused on those that they are unable to match those to the actual needs of stakeholders. So they produce evaluations and they check the boxes and are able to say yes we did it. (IP16)

The federal policy informs the standard provincial policy, which in turn informs the local and organizational policies and then informs the funders of course and their role in that whole mix. If you have this really rigid policy, if it's too rigid or too prescriptive at the federal level, then you might not get good evaluation work happening. On the other hand, if it's not prescriptive enough or it's too "loosey goosey" where there's just nothing there, it could have exactly the same effect in terms of hindrance that a too prescriptive policy would have. (IP5)

My findings also suggest that overly detailed evaluation policies not only restrain capacity and are difficult to maintain but also are particularly perilous for the quality of evaluation and thus may do more harm than good. As one of the participants pointed out,

In my role as a reviewer of evaluation for the EU in the field of regional development and crime and justice I see this happen. The EU is prescribing its evaluation policy and the rats pop out of the earth, not bureaucrats but rats. Folks that have no idea what they are doing but they use the right words and then they say "Yes we do want evaluation policy" and then they earn a lot of money and they deliver really low quality evaluations because the evaluation policy is prescribed and articulated and they have small little booklets describing the words. These guys are copying the words and putting them out to tender and then they get it done. This is really a perverse effect of evaluation policies and I see it happens all the time. (IP2)

Nonetheless, many participants agree that effective evaluation policy is both a source of information and a tool for highlighting potential areas of concern and for offering guidance to approach these areas without restricting evaluation practice. When there are restrictions on evaluation practice, capacity is limited. For example,

If you want to improve practice and ultimately use, policy could help but you don't start off saying evaluations should be of high quality and use high quality processes and methods and then jump in to say all evaluations therefore need to be randomized controlled trials because you think that's high quality. This is a kind of prescriptive policy that really constraints people who are actually high-performance practitioners. (IP3)

There is widespread agreement among the evaluation scholars and practitioners I interviewed that the contextual factors embed the role of evaluation policy in building organizational capacity within a broader, interconnected system that is influenced by socio-

political, economic and historical factors. Legislation, government policies, national economic conditions, culture and global influences are all contextual factors that moderate the role of evaluation policy in building organizational capacity for evaluation and how people perceive evaluation. The content of the evaluation policy itself is also a significant macro-level moderating variable that shapes evaluation practice and capacity-building processes, especially when a policy is externally imposed upon an organization. Thus, recognizing the effect of these contextual factors at the macro level is critical for understanding the dynamics of the role of evaluation policy in building evaluation capacity and for identifying the challenges involved in the process.

Meso Level: Organizational Context

The meso level in the ecological framework refers to factors within the organization that influence the effectiveness of evaluation policy and enable or hinder its role in building capacity to do and to use evaluation. It is at this level that the interplay between the macro level (socio-political context and evaluation policy) and the micro level (stakeholders) takes place. The data show that the moderating factors at the macro level have varying influences on the capacity to do and use evaluation at the organizational level.

Capacity to do evaluation. The findings reveal that the impact of evaluation policy on the capacity to do evaluation at the organizational level is affected by the existence of a supportive infrastructure that includes the availability of resources, the commitment of leadership and the presence of an evaluative culture that supports evaluation.

Most of the interview participants shared the view that the existence of a **supportive evaluation system** within the organization is essential for setting the stage for evaluation policy fostering evaluation capacity building activities and contributing to the improvement of the state of evaluation within the organization in general, as is suggested by the following comments:

Evaluation policy can light a flame but the flames are only going to catch it if there's dry timber and so a certain infrastructure has to be in place to see an impact. (IP16)

I think evaluation policy is a wonderful tool but I don't think it's enough. What the French say is "La plus belle fille du monde ne peut donner que ce qu'elle a," meaning that there are only some things that you can do. You need a lot of help. You need resources, you need a system, you need big support from management. That's really important. (IP11)

An essential aspect of the organization infrastructure is the degree of coordination within the system to ensure access to evaluation resources and information. This is affected by the alignment of work processes and by the nature of data sharing between units and departments. For example, some participants shared that it is mostly the case that the collection of evaluation data is the responsibility of another unit within the organization and that, without the proper coordination, the potential of evaluation policy to streamline evaluation capacity building activities remains unrealized. This is particularly the case with the ongoing collection of performance measurement data by program staff, and its subsequent suitability to be used for evaluation. As one participant explained:

The policy should refer to a specific systems and structures and resources. Those are very important things to be in place when the policy is launched rather than to say, well, the policy said in five years you need to be this. Well we are talking about performance measurement here. It is a problem for every organization. Performance measurement doesn't feed into evaluation enough and the problem is the quality of the data and the availability of the data. Performance measurement is always left to the programs and then evaluation walks in and says, well, we are going to use your performance data for the evaluation and then they find out it doesn't exist. I think those things have to be in place. Otherwise there is not even a point of having an evaluation policy because there is nothing that you can really use. (IP14)

Another important variable that moderates the role played by evaluation policy is the availability of appropriate **resources** to effectively implement the evaluation policy and to turn its strategies into actions that are able to accomplish its objectives and, ultimately, achieve the desired impact in terms of organizational capacity for evaluation. This includes the provision of resources to support staff participation in evaluation capacity building activities and the allocation of sufficient time to integrate the policy into the organizational system. As some participants pointed out,

Give enough money, serious money, real money and then give it time. It's really difficult to expect a serious result for anything in a one-year or two-year period. Give the policy enough time in the real world to make its mistakes and learn from them and get it back together. (IP7)

Organization capacity is based on its resources, but the resources don't follow from the policy. It could very well be that the policy might affect where the evaluation fits in the organization and who it reports to and its authority to define evaluations. It could also insist that there is cooperation in evaluation, but without resources how are you going to do that? (IP6)

Some participants also noted that resources are needed in order to bring organization members on board. They must have enough time to implement any additional evaluation

activities that they are not currently performing. Otherwise, they may resist becoming involved in the capacity-building process. Indeed,

If a policy is budgeted and funded it will make difference. Otherwise it will only add extra work to people who already have a job to do. You can't expect people to do things and appreciate them if you only ask them to work harder! You have to make the benefit obvious and transparent to convince people. To do so you need time and money. (IP9)

Other participants noted that people in general are interested in capacity building and professional development activities for various reasons including learning and improving their skills and competencies as well as enhancing their career progression. When evaluation policy promotes professional development, people will pay more attention to it and will become interested in it, though there must be sufficient time and resources to allow people to interact, collaborate, and participate in such activities. One participant provided an example that explains this point:

Time, money, interests, collaboration would be the four things that immediately came to my mind. And the lack of these factors will hinder benefiting from evaluation policy. I will give you an example. For the last decade a group of colleagues and myself have run the Irish evaluation network which is an organization that seeks to bring together evaluators in Ireland in an informal way four to five times a year. We put workshops and seminars around different themes in evaluation. We run that because I work in the university and I have a little bit of research funding and I can do that. We don't have lots of money and we don't have lots of time and there is quite an interest in a lot of different sources about these things but it never goes beyond really just seminars and online distributions of those seminars. If we had more money and if we had more time, we would be able to do a lot more, so I think these are critical things. (IP15)

The findings suggest that, in order to ensure that evaluation policy has a positive impact on evaluation practice and promotes capacity-building activities, the organizational leadership must support evaluation and must communicate its importance to improve the organization's effectiveness to both staff and partners. Leadership support is necessary and can make a significant difference in terms of establishing an organizational environment for ongoing learning and improvement in which organization members want to consistently perform at their best, as these participants explained:

Unless evaluation and capacity development are supported by senior managers, they won't be prioritized throughout the organization. Policy could help with that but you need continuous support for learning. (IP18)

You need to have organizational leaders who recognize the value of having an evaluation policy and who are willing themselves to talk about evaluation policy and contribute to its development and that is a big part of it. (IP4)

Thus, the right leadership can breathe life into evaluation policy and can move it from a static document to actions that boost the opportunities for learning and capacity building. In the following excerpt we can see how leadership and resource provision may be interconnected moderating variables.

Resources and leadership are the two important components for evaluation policy. The policy cannot be effective if the evaluation budget is one percent. Sufficient budget could be the key to successful evaluations or the reason for the failure of evaluations. Evaluation policy can outline this issue. What I want to say is that evaluation policy will do nothing if the leadership or the management do not act on it and support it. And you can flip this around. (IP17)

Therefore, for evaluation policy to impact capacity and foster the integration of evaluation into the ways in which organizations do their work, it is critical that organizational leaders believe that evaluation is an important and meaningful practice. They have to consistently communicate the importance of evaluation for decision making and to encourage organization members to engage in evaluation activities by providing the time and financial resources necessary for ensuring that evaluations can be conducted. Many participants noted that, for this to happen, there must be a true commitment to evaluation because, even though resources are essential for evaluation capacity to evolve, it takes commitment across the organization to elevate evaluation practices and their use and to ensure that these practices enable learning.

Many participants shared the perspective that evaluation policy's impact in terms of fostering capacity hinges on the extent to which the organization is committed to evaluation as a vehicle for learning and change. In other words, evaluation policy will have little to no effect on capacity building if the organization lacks an evaluative culture that supports making evaluation strategic and useful, which in many cases is established at the management level. One participant even argued that, when organizations are committed to evaluation, limited resources are not be an obstacle for evaluation; instead it is a matter of the management being strategic with the available resources and being more intentional about learning and evaluation activities. As this participant explained,

I believe a commitment to the broad idea of evaluation as an integral part of how an organization or a system evolves or learns has to be there. If you don't have that broad level commitment, then I think all is lost. Other issues like resources become an issue only if an organization does not have a commitment to learning. If an organization is committed to learning, it needs information and it needs feedback and whether we are talking about performance measurements or we are talking about the deeper evaluation that looks at the nature of what you are doing, how well it has been carried out, how effective it is and so on. (IP13)

Again, we can see the interconnection among meso-level moderating variables, in this case evaluative culture and resources. Another participant also emphasized that:

Organizational culture is very important to support the evaluation function. Allocating resources to evaluation is important to conduct and use evaluation as well as developing the capacity of the staff to do training. I think organizational culture and resources are related and both of them are fundamental factors for capacity building but I think the organizational culture comes first because nothing will happen if the organization doesn't believe in evaluation. (IP1)

Thus, the findings reveal that organizational culture is a key meso-level moderating variable that continues to shape evaluation capacity and the role that evaluation policy may play in enhancing evaluation practice and use. According to most of the participants, organizations must have a culture that encourages learning, risk taking, and dialogue to improve their performance if they are going to benefit from a well-thought evaluation policy to create or support a dynamic environment. As one of the participants explained,

You need the right kind of organizational culture. I wouldn't even say an evaluation culture but sort of a learning-oriented culture centred upon improvement where staff and managers are rewarded, not for meeting objectives but for asking difficult questions, for experimenting or trying different things even if they don't work. That's really what it is. People can then see that evaluation is a tool to help learn how to do better. If they really believe in learning and improvement and they can reinforce that with the evaluation policy then they'll see it as a tool. (IP17)

Others participants stressed the importance of the existence of an organizational culture that ensures that there is no resistance from organization members towards the evaluation policy and that the policy can be successfully integrated. As one participant noted:

If you want to put together an evaluation policy, you have to be able to do evaluation. If you do not have that I think that will hinder the success of the policy because you want to be able to get it off the ground. I think it is about organizational resistance and organizational culture. Organizational culture can support evaluation policy and organizational resistance can hinder it and again you can't get it off the ground because there is so much opposition to it. (IP4)

A number of participants also noted that the ability of evaluation policy to influence the capacity to do evaluation also depends on the type of organization, as this affects the evaluation culture and how organization members perceive evaluation. Indeed,

I think the degree of hierarchy and formalization of tasks within the organization will have an important impact on the success of the evaluation policy. Highly formalized organizations may be more receptive because they have policies for everything else, so this is one more thing that help them structure their work and identify roles and responsibilities. In more open organizations where things are not formalized, like small community organizations, people may perceive evaluation policy as being very threatening. So I think culture is a really big part of that and structure as well. (IP1)

Many participants also shared the view that the type of organization and the stability of its management and governance structure affect its culture and influence both the role of policy in promoting evaluation capacity and the feasibility and sustainability of capacity-building efforts within the organization. Two of the participants noted that this is mostly the case in government organizations:

I don't think it is feasible for evaluation policy to influence organizations that are run by governments because evaluation people come and go. If one of them comes in and sets evaluation policy to focus on learning and capacity building, I do think sustaining that after that person leaves is a real challenge. For example, urban governors have come in and set evaluation as part of their office. My experience has been those offices never survive in changes of administration because it is politically viewed as part of the last governor, so the new governor is like okay I got my own stuff so these are washed out without even thinking about it. We deal with this all the time. (IP16)

Evaluation policy will inform the big picture. In terms of a formal evaluation policy that is really thinking about the big picture and how to build a learning system cycle, you are not going to find a lot in public sector. I'm not aware of any state that sort of has strategically what we want to be learning as we are doing things. (IP3)

Notwithstanding the type of organization and the importance of stable management and a supportive organizational culture for the evaluation policy to effectively influence capacity, evaluation policy can be a part of evaluation capacity building efforts within the organization, as one participant argued:

I would argue that if you have for example an evaluation function and there are three or four people in it, whether they're doing evaluations or contracting it out or doing a mix of both, it doesn't really matter but all of a sudden they all leave. You should be able to turn around and replace them and the new staff should be able to come into something without starting from scratch. There should be a policy. A policy is part of an organizational capacity. There should be a policy, there should be a set of standards, there should be processes, there should be procedures. There are all sorts of stuff that need to take course over time that are adapted or

evolved in relative sync. Ultimately an evaluation policy is part of organization capacity for evaluation. (IP15)

It is interesting to note that this participant appears to be arguing the merits of bureaucratic formalization and its direct implications for evaluation policy's role in ECB. This underscores the importance of the existing evaluation culture and system within the organization and the fact that the organizational capacity for evaluation is an "ever-evolving" process that happens over time and that evaluation policy is part of such a system that influences and is influenced by the process.

Capacity to use evaluation

The findings reveal that the impact of evaluation policy on the capacity to use evaluation at the organizational level is moderated by the integration of evaluation into decision-making processes and by the conceptualization of the purpose of the evaluation function within the organization.

Many of the interview participants stated that the **decision-making process** within the organization **and the role of evaluation** within this process affect the role of policy in developing the capacity to use evaluation. They shared the view that the ability of evaluation policy to positively foster evaluation use depends on the extent to which the management of the organization relies on evaluative information and uses it for decision making. Indeed,

I think evaluation use mostly depends on how managers make their decisions and where they get the information they need. Do they get it from evaluation? They don't need a policy to do that but the policy may streamline the process. (IP18)

A policy is just a statement. It may mandate certain things to take place but it can't force people to think in those terms and to internalize it. Managers are going to use evaluation because they need it, they believe in it and they understand how it is valuable to them, but this is not just going to come through a policy. There may be a policy that for example makes evaluations take place and maybe the policy does focus on use. It could be a greater simulation. But if managers see that the use will be negative they may really try to avoid it and that often happens. (IP9)

Participants further argued that, despite the existence of an evaluation policy, the significance and usability of evaluation information would be limited if evaluation is disconnected from organizational-level decision making. If management does not demonstrate to the staff how they can integrate evaluation and use it to make decisions, to take action and to learn, the potential of evaluation policy to enhance evaluation use remains unrealized, as is implied by the following comments:

It's not just the policy that is going to make that difference in itself. It's one way, but we're talking about evaluation use. Does top management live the talk? Do they really talk of evaluation, can they give positive examples of how they use evaluation for their decisions, how they reinforce and reward people who have done evaluation and been subject to evaluation? That's one thing. Do they really take it seriously themselves. If it is in the policy but you never hear about it from top management, then other managers are not going to take it seriously themselves. (IP14)

A lot of evaluation use has to do with management. I would sort of agree that, yes, policies make things explicit and I can see some linkages between policy and whatever its people are doing, but I don't think that it will make a big difference in the ability to use evaluation. What I'm saying is that even with the best implementation it's not going to compel use. (IP7)

Some participants also pointed out that, even though there is an evaluation policy at the government level that demands evaluation, evaluation is a fill-in-the-form and check-the-box exercise that is viewed as a burden and as having no impact on building organizational capacity because it is not integrated or used for strategic decision-making. As these participants explained,

To a great extent in the federal government for sure evaluation is done to check the box and that is the only reason they do evaluation. There is no intention of using it ever I don't think, not for strategic decision making. It is used for program improvement to make minor adjustments but federal evaluations never say this program is not achieving its objectives so we need to get rid of it. They never say that. They never help high-level decision-making but they do support anecdotal changes to programs. So I think evaluation policy can be very helpful but it can also become another requirement that people are expected to meet and they would check off the boxes without any kind of real capacity to use evaluation. (IP10)

Evaluation is to be taken place by decision making. Some organizations say decision making should take into consideration the environment, consider human rights, consider this, consider that and consider tick box things which don't really necessarily change very much but the formality of it. If you really want evaluation to be used, you need to get at really the mentality and where people feel there is a point to it or it's useless to them. Otherwise it's just a form. (IP2)

This could be attributed to the fact that the mentality around evaluation use in most government organizations is more summative or accountability-focused than formative. While formative evaluation focuses on learning and serves the needs of managers and organization members who are interested in improving their programs, summative evaluation focuses on accountability and is mostly intended to serve decision makers who are concerned with decisions about the distribution of resources; therefore, it is perceived as a control mechanism rather than as a learning and improvement process. In such cases, the role of

evaluation policy in building evaluation capacity is compromised. Instead, evaluation policy is most likely perceived as a mere tool for enforcing control. Indeed,

I'm not aware of any state that sort of has strategically what we want to be learning as we are doing things, because this is not going to be the first thing they think about. They think about accountability, they think about political aims and they think about management. It is difficult for government to focus on capacity building and learning and these sorts of things. On the average, policy makers in the executive branch of the government are typically there for 18 months to three years, so policy makers and legislative folks come and go. So it is tough to keep that capacity to the forefront. (IP11)

Some participants also explained that the prevalence of the symbolic use of evaluation as a strategy for retrospectively justifying decisions already made within an organization not only limits the role of evaluation policy in fostering evaluation capacity but also reflects the persistence of an unfavourable gap between evaluation practice, on the one hand and the use of evaluation as a mechanism for learning and improvement, on the other hand.

Alternatively, the evaluation follows after the decisions have been made without any management intention of utilizing its findings. According to one participant,

I think this is the challenge of any evaluation is how to get the evaluation used. Part of the problem is in Europe—I'm not sure about Canada or other parts of the world—is that evaluations are used to prove success not necessarily to come up with judgment of value. So if you do evaluation as an approving exercise, then there is little or not enough reason to utilize it afterwards because it already told you what you want to learn in the first place, so what is the point of having it. We have all kinds of policies but I'm not sure how they would fix this problem. (IP15)

In this sense, many participants expressed concerns that evaluation and evaluation policy are used as tools to preserve the status quo rather than to explore assumptions.

On the other hand, some participants provided examples where the evaluation policy was acted upon to enforce evaluation use. The examples show that, when evaluation information is clearly integrated and connected to decision-making processes, the evaluation policy is can be used to signal to staff and the organization what the evaluation is going to be used for and on what values the evaluation practice is based. One participant shared the view that:

In the military there's an extremely specific evaluation policy for personnel. The military not only has very detailed criteria. They have extensive training in capacity building that's tightly linked to the use. In all of the branches there are pages after pages of policy defining the performance expected and how the performance is to be documented. Nobody gets promoted

unless these evaluations are done. Here's an instance in which evaluation policy and evaluation use are glued together. Everyone takes it seriously. (IP12)

While this example speaks more to personnel evaluation than to program or intervention evaluation it does emphasize the point that evaluation findings are more likely to be used where they are meaningful and significant to the implicated actors. Another participant shared an example that shows how evaluation policy influenced the capacity to use evaluation by intentionally integrating evaluation into the decision-making process:

In the federal government it was really felt that evaluation was not on the radar of senior decision makers. They barley knew it existed. They knew that they have it but they did not know what it is and they never saw the reports. So evaluation policy said there is a huge gap in terms of the capacity to use evaluation because it is not making its way up to senior decision makers. So we are going to make that mandatory. So the policy spells out that we are going to have a departmental evaluation committee. This committee is going to presided over by the deputy minister and the deputy minster needs to personally approve of the evaluation reports which ensures that he or she reads the reports. And now several years later you have deputy ministers who understand and know about evaluation because they have been exposed to it for five years. (IP16)

One of the participants further emphasized the following:

One way to answer the question of whether an evaluation policy is positive in terms of evaluation use I would say, yes, inasmuch as it is integrated and absorbed of manner by the evaluation function. If it is a living document or framework, not something you do once, then set it aside and never think or worry about it again. When it's something that's integrated then the probability of it being positive it's obviously much higher than if it's just a book or a document that goes on a shelf and nobody looks at it or cares about it until the next time it's reviewed in ten years. (IP1)

Again, the meaningfulness of the information evaluation can provide appears to be essential to its propensity for use.

The second key variable that moderates the impact of evaluation policy on the capacity to use evaluation at the organizational level is the conceptualization of the **purpose of the evaluation function** within the organization. The findings reveal that the purpose of evaluation activities within the organization and what organization members *perceive evaluation to be* is critical for the role of evaluation policy in fostering evaluation use. Indeed,

It depends on the extent to which what you conceive to be good evaluation and good evaluation practice and values lines up with whatever the policy thinks it is. An evaluation policy that says evaluation is a tool to assess the extent to which programs have achieved their objective and then to report on these for accountability purposes including payment and

budgets and so on and so forth. Now, if that's how the function or the individuals who are carrying out evaluations, if that's what they consider evaluation to be and that lines up with their understanding, their vision and their conception of evolution, then obviously that's not a problem. If on the other hand a policy is different from or it's an opposition to what an individual or a group or a function would hold to be valued in evaluation, then clearly it's problematic. (IP9)

Evaluation is actually something that means something when you're tasked with either carrying out evaluations or contracting them out or managing them. (IP17)

The findings reveal that the common understanding of the meaning of evaluation within the organization and the familiarity of organization members with evaluation principles moderate the extent to which evaluation policy is going to have an impact on the use of evaluation. As one of the participants explained:

The term "evaluation," people have different ideas about what that means. I realized that this is why people are having trouble understanding what evaluation policy is. So, there are different ways of viewing evaluation and then there are different ways of understanding what policy is. As a strategy tool to guide and direct how the ways in which evaluation is used, it depends on how it's conceived, how it's looked at, how it's conceptualized within the organization and how that's used to guide and direct evaluation practices. (IP5)

Many participants also commented that, quite often, organization members, and sometimes managers, confuse or conflate evaluation and audit as a single concept or function within the organization. This confusion undermines support for and use of evaluation as a mechanism for learning and improvement and, ultimately, limits the role of evaluation policy. For example,

Evaluation policies that would say to check the extent to which objectives have been achieved and report and make decisions on funding, for me that's not evaluation. For me that's audit. And if I were told you have to do evaluations and follow this policy, I wouldn't. (IP18)

Another participant also explained that

Organizations that are only conducting evaluation to meet regulatory inspection people are still not seeing evaluation as useful or as something that has benefits to them. They see it as a waste of time and conducting evaluation is a burden. So this definitely influences how it is used. (IP9)

While the purpose of evaluation is to understand and assess the merit and worth of a program and to identify the best alternatives to inform decision-making and generate improvement, the purpose of auditing is to assess the degree of compliance of a program with predetermined standards or objectives. When evaluation is done in an audit-like fashion, it

loses its meaning, value, and purpose; the knowledge that the evaluation eventually produces is either useless or irrelevant for capacity building and improvement activities. As one participant pointed out,

Evaluators deboned the fact that auditors are encroaching on their territory when in fact it's the evaluators that are set out to occupy the accountability territory, which was theirs to begin with. This to me is sort of the broader context and of course in all of that, you've got tons of people running around that haven't got a clue. Either carrying out or purporting to be carrying out evaluations, leading evaluation units. So the question of the value of evaluation can be a very topical one right now. The extent to which that connects to the evaluation capacity to conduct and use evaluation, I think, is key because it comes back down to, what is the value of evaluation? (IP8)

It should be noted that ECB, as defined above, refers to *intentional and directed* efforts undertaken by the organization to become more reflective and more involved in evaluation processes. If the purpose of the evaluation is blurred or unclear, evaluation capacity is affected because it is challenging to make intentional efforts when there is a lack of clear purpose.

For many participants, the primary focus on accountability as the major purpose of evaluation has significant implications for evaluation, especially in terms of its learning benefits and use. One of the participants shared an example to explain how the exclusive focus on accountability in evaluation policy caused significant harm instead of fostering use:

In Europe, in the evaluation of education at the moment, the question is how strong the accountability element in the policy should be and what are the unintended consequences that the focus on accountability is causing damage. Well, the stronger the accountability element is the more likely people are to resist the system by for example changing exam results. There are teachers who are sent to present for changing test scores because of how high stakes the evaluation is. On the other hand, people do respect and expect a reasonable degree of accountability in the policy, so it is a tension. (IP14)

Another participant also added:

If organizations have a lot of fear of accountability, oh you know, I've got to basically cover my ass therefore I'm going to use evaluation as a way to do that. It's more of a sense of I have to become more protective, I have to protect myself or my organization or whatever versus seeing it as a valuable resource as a positive, proactive, value-added type of thing. It's very different and the policy you have set, if it's informed from the fear perspective, it's going to be very different from the value-added perspective. (IP17)

Nevertheless, a number of participants acknowledge that accountability is an integral part of evaluation and that evaluation policy should reflect this reality. Yet, for evaluation policy to

foster evaluation use, it should make accountability meaningful in a way that reflects the true value of evaluation. As one of the participants explained,

The purpose of policy is accountability but accountability could mean different things. It could mean did we meet all of our goals, which is really low bar. But evaluation policy could make accountability more meaningful, like what were the needs of our service recipients and were we able to meet their needs? No matter what we're doing we want people that are service recipients to be doing better. We want to enhance their well-being. We want whatever's going on with them and whatever services we are providing to benefit them. That's where accountability is. It is to understand what our services are doing and in what ways we are benefitting or not benefitting our recipients. If the evaluation policy is being set in such a way and articulated in a particular way then accountability becomes more meaningful. (IP12)

Many participants also stated that, when evaluation is mandated by external factors, especially in order to secure funding, organizations comply with the requirements and evaluation is conducted only to meet those specific requirements. In such organizations, evaluation is not perceived as bringing benefits to the organization other than funding, as one of the participants explained:

My point is that you have to convince people and make the point that evaluation is a good thing and it is useful. Not sure if the policy will make a difference. If you can't show people that evaluation can benefit them and benefit the organization, then why did you ask them to do it? It doesn't matter if you have a policy or not. You know what? One time I was presenting the findings of an evaluation that I had done and one of the people in the organization said we got the funding so just move on. It was terrible. I said you don't want to learn about it and he said it's okay we got the funds we need. (IP9)

This rigid framing of the purpose of evaluation as simply a tool to justify continued or discontinued funding for a program undermines its use. Moreover,

I am so aware of how people are to make themselves look good. All they want is the funding and they'll put their best foot forward but when you see what's actually happening on the ground, it's much more hit or miss. (IP7)

Some participants commented that the evaluation policy is likely to make a difference in terms of how evaluation is perceived and used within the organization when the policy both reflects and is supported by the organization's mission and values. Indeed,

If the policy is well thought out it would help the organization internalize the value of evaluation for their own purposes and to improve their programs not only to respond to funders' requirements. Evaluation policy should be connected to the mission statement of the organization. It would help the capacity and it would be another tool that helps support and advance the mission and the goal of the organization. (IP11)

Another participant provided the following example of how the evaluation policy influences evaluation capacity when it is clearly connected to the organization's values:

So back to the question of how well is the relation between the policy and the evaluation practice and use? The answer is yes. It seems like a reasonable assumption but the policy should show and explain how the values of the organization are woven into the way evaluation is conceived and into the way in which it's practised and used. If, for example, a value of an organization is evaluative transparency then you would expect that in the evaluation policy that particular value would be woven into the policy. As, for example, all evaluation reports will be made public. Now there is a relation. (IP8)

It should be noted that the majority of the participants were of the opinion that, for the evaluation policy to leverage the capacity to use evaluation, there should be implications for failure to comply with the policy. The role of following up on evaluations is emphasized in the following quotation.

My actions have consequences. No consequences, no evaluation, no use. That's no feel-good information but it's honestly what I think. When you have an evaluation policy in place that outlines what evaluation looks like, what resources are available and who is responsible then people would read it and say, okay, well that's good. Without follow up, without consequences, without working on it to integrate this practice or the element that the evaluation policy has within the evaluation practice, then nothing can happen. (IP4)

On the same issue others stressed the need to include enforcement actions and strategies as an integral part of the evaluation policy in order to enforce its role in building organizational capacity. According to one of the participants,

Evaluation policy has to have some teeth. There has to be consequences for not meeting the requirement of the policy. An organization really struggles with this but evaluation is one more thing sitting on the desk of a program manager where he has too much on their plate. So, if there isn't a real consequence of not doing something, it is not going to get done. And I totally understand that. I'm the same way and everybody is the same way. I get it. So if you are telling program managers that evaluation policy is a priority, there has to be an incentive or there has to be a consequence because your policy has to have some sort of teeth to it. Otherwise there is just probably no point. (IP10)

One of the participants described the need for enforcement processes to ensure use as an important aspect of the evaluation policy in order to foster capacity, especially in organizations where evaluation is outsourced:

The evaluation policy should include direction about use. This is very important. Our policy doesn't have anything about the use of findings. An external evaluator is appointed for three years and in the last year he writes a report and that's it. We don't have any contact with that evaluator and they are not going to follow up with you. Nobody will ask you if you act upon

the recommendations. So there is no system for compliance with recommendations and that's probably because it isn't written in the policy. (IP9)

Another participant provided an example of an enforcement strategy that uses a rating system to ensure compliance with the requirements of the evaluation policy. The participant argued that such a strategy has strengthened the evaluation capacity of the organization. As the interviewee explained,

In the federal government the conformity to the policy or the requirements of the policy was assessed for ten years annually by the Treasury Board Secretariat through the Management Accountability Framework. So they would assess every element of the policy and the extent to which it was implemented in each of the organizations. If the organization got a bad grade, the performance pay of the deputy minister would be affected. So the performance pay would be more important to some people than others but because it was a rating and a public grade that was given to the function, all of the sudden everybody wanted to do well. So those are incentives to meet the requirement of the policy and develop the capacity. (IP16)

Thus, there is a broad consensus among the interview participants that the role of evaluation policy in leveraging capacity to use evaluation at the organization level is moderated by a number of aspects related to the extent to which evaluation is integrated into decision-making processes as well as to the perception of the evaluation purpose within the organization. For evaluative information to be used within the organization, it has been routinely suggested that it has to be meaningful, of some learning value and perhaps through some combination of support and pressure, be acted upon. The findings reveal also how macro-level variables such as evaluation being imposed on the organization, influence, and shape variables at the meso-level, such as perceptions about evaluation being exclusively accountability-oriented. Also noted are interconnections between meso-level moderating variables associated with capacity to do evaluation and the capacity to use it, leadership, and integration of evaluation into decision-making being a salient example.

Micro level: Evaluation Stakeholders

The micro level refers to the evaluation stakeholders within an organization. The analysis reveals that the above-discussed interconnected moderating variables at the macro level (context) and the meso level (organizational) have different influences on the role of evaluation policy in building capacity to do and use evaluation at the micro level (individual organizational stakeholders). In terms of the capacity to do evaluation, I found the role of evaluation policy in influencing capacity building at the micro level to be moderated by two

variables: the evaluator's role and skills and the interpersonal relations. Evaluation policy's influence on the capacity to use evaluation is moderated by the stakeholders' engagement in evaluation and learning.

Capacity to do evaluation

The findings show that the role played by evaluators within the organization and their evaluation knowledge and skills influence the effectiveness of evaluation policy and enable or hinder its role in building evaluation capacity. The technical and **interpersonal skills** required by evaluators to enhance the role of evaluation policy are related to the identification of evaluation issues, the use of appropriate data-collection methods, and the generation of evidence-based recommendations. Softer skills such as building client trust, communicating evaluation messages in a clear and transparent way, and meeting stakeholders' information needs are among the communications and interpersonal skills evaluators use.

Many interview participants shared the perspective that having evaluators with adequate skills and training to design and conduct evaluations as well as to plan for and facilitate learning from the evaluation is a key for ensuring the ability to act upon the evaluation policy. As one participant explained,

If you have a good evaluation policy and a professional evaluator, I think yes evaluation policy would be extremely influential in terms of influencing practice. It will make evaluation visible and legitimate, but you need skilled evaluators who can work with people to develop their evaluation capacity. You need evaluators who have high communication skills to raise the importance of evaluation beyond the evaluation unit. (IP4)

Other participants also pointed out that evaluation policy can set the stage for evaluators to play a key role in educating organization members and in improving their knowledge and skills regarding evaluation. For example,

I think policy is important in terms of setting the stage and stating what's important, but if I'm not an evaluator and I go okay we're supposed to do this and this now, what do I do? So, I think people need to understand evaluation and why it is important. They need to be educated. I think evaluators should take that lead. (IP13)

Some participants also mentioned that evaluators could use evaluations as opportunities to inform organization members about the role of evaluation policy and to demonstrate to them

the value of evaluation as an instrument for learning and improvement within the organization. One participant commented that

The best way to get these policies known is when something has to be evaluated. I think the evaluators are the best people to talk about the policy. They are the passionate people to talk about it. They are able to answer questions and explain things so they are best placed to do that. (IP18)

One participant further argued that, for the evaluation policy to help improve organizational capacity, evaluators could act like coaches to guide organization members to achieve the goals of the evaluation policy and to support them in building their capacity and knowledge about evaluation. Indeed,

When the organizations have a policy or guidelines I think the evaluator needs to take on the role of a coach, coaching people about evaluation to keep them moving along. The last thing you want is this whole document to sit on the shelf because people don't know what to do with it. (IP5)

Another participant also emphasized the point that evaluators need to play an active role in facilitating understanding of evaluation by using their skills and knowledge to discuss and clarify the assumptions embedded in the evaluation policy. In this participant's view,

I think policy and capacity are kind of like the mind and the heart. I think evaluators have to change attitudes by improving awareness and helping people understand. People have to understand what we mean by evaluation. We have to use the same language and the same definitions. Everybody in the organization has a different idea of what evaluation means. Never do we all come to the table with the same understanding. So, let's talk about what it means and here is the policy to help us. (IP17)

For evaluation policy to influence organizational capacity for evaluation, some participants further explained that, evaluators should also work with the management to promote the value of evaluation as a management tool. As one participant argued,

the directors of the organizations are not trained in evaluation and they don't have the skills, so they do the best they can to be compliant but they don't have the background or training like evaluators. They are not evaluators. Evaluators need to work with them to get there. (IP7)

Another aspect of role was associated with communication. Some participants argued that the evaluator's communication skills are critical for promoting evaluation and for using the policy to foster evaluation capacity. One of the participants commented that

people need to be pulled into evaluation more than pushed to evaluation. You can push them to evaluation through mandates and policies but I don't think that's going to get you much more than sort of check-off-the-box answers. (IP1)

Therefore, communication between evaluators and other organization members is an important factor that affects not only whether or not they are supportive of the evaluation but also their perception of the value of the evaluation and its use, as one participant explained:

People will use it or they won't use it when it's in their interest. The question is how can you persuade them that it's in their interest. I think that's up to the evaluator. I don't think that it's up to the agency. You need evaluators who know how to communicate and negotiate to persuade people. I don't think a policy is going to get agencies ever to use evaluation work that is not in their interest.... It's a tension, it's difficult, but it can be done if evaluators know how to negotiate. Nations negotiate. I mean we're negotiating with Iran for God's sake. (P11)

Thus, the role that the evaluator plays within the organization moderates whether or not the evaluation policy is going to have an impact on organizational evaluation capacity.

Evaluators with the right skill set, and perhaps especially soft skills, play a critical role in helping organization members to understand the value of evaluation and to develop their knowledge to address the long-term change that the policy is aiming to achieve.

Capacity to use evaluation.

The findings reveal that the impact of evaluation policy on the capacity to use evaluation at the individual level is moderated by the engagement of stakeholders in the policy-making process and/or in the evaluation process and by the learning that takes place during and after the processes. The vast majority of the participants were of the view that, for evaluation policy to improve evaluation capacity and ultimately evaluation use, **stakeholders must be engaged in the process** of developing the policy. They emphasized the importance of ensuring the responsiveness of evaluation policy by actively involving those individuals who will be affected by it, as is implied by the following comments:

The more individuals that are involved in a process, the greater the likelihood that they'll find the process and some of its critical results such as the actual document credible, meaningful, and so on and so forth. The effect of an evaluation policy on evaluation capacity is also a function of the process by which the policy was developed. They taint and ultimately integrate to which the function plays out. (IP6)

I think the policy can increase the function of evaluation as a learning tool if the policy is worked out with the organization. What do you need to learn from the evaluation? The policy

must be discussed and agreed on by the people. There must be participation in creating the policy and willingness to use the policy. (IP13)

There is no one size fits all when it comes to policies. Evaluation policy should be tailored and based on the expectations of the organization, and doing so requires stakeholders' engagement so it resonates with them and their needs and with the decision makers. (IP10)

The participant quoted hereunder provided a vivid example of a participatory policy-making process at the government level. The respondent explained the ways in which the engagement of stakeholders in the policy development process contributed to their knowledge, enhanced their understanding of evaluation, and brought into the policy discussion the expertise, perspectives, and ideas for alternative actions of those concerned. As this participant explained,

I drafted an evaluation policy for the government of Denmark that they took and we worked on it and we presented it to the committee of parliament and we discussed and debated and eventually came out with an evaluation policy. The crafting, did it help? Did it help practice and use? I would say first of all, just the fact that the process of crafting the policy involved a range of different stakeholders including a lot of stakeholders who weren't into evaluation and a herd of them didn't know what it was about, then as a result of the process they ended up knowing about evaluation and what it was for. In and of itself it is positive in my view because people's knowledge is increasing and enhancing and it puts them in a position where they can participate in evaluation-related discussions and debates and put them in a position where they're more informed than they were. They didn't know about this and it's by virtue of the action process, discussion, debate it creates its own values, so to speak. That I view as capacity building. (IP15)

Stakeholder engagement can also enhance voluntary compliance with the policy for two reasons. First, stakeholders may adjust to changes more easily because they are made with opportunities for input to allow stakeholders to anticipate and overcome challenges in a timely manner. Second, stakeholder engagement engenders a sense of legitimacy and shared ownership that motivates them to comply with the policy.

Many participants also commented that stakeholders' active engagement in the policy-making process is more effective than the provision of these individuals with information about the policy through less engaging vehicles like seminars. For example,

If you have a training session on [the evaluation policy] next week and you have to go to that and read everything else and people go to that and say oh god I have to go to this policy training session and yet another policy and listen and it just doesn't make any sense to them. If you are on the other hand involved in the development of it then it is participatory. They start to appreciate and internalize how evaluation might actually be helpful to them. (IP8)

Similarly, another participant reflected that

The more participatory the policy is the better it will work. It might be slower, but it will be more effective. Involving people in every level of the policy, developing it, implementing it and so on. People need to know how to use the evaluation policy properly. You can't just put it over them. (IP1)

On the other hand, one of the participants expressed concern that there be a balance in the level of engagement of stakeholders in developing evaluation policy in order to maintain the credibility of the policy and the quality of its results:

There is a lot to be said for having everybody have a vote in this because generally speaking you don't distrust anything you have been working on yourself. On the other hand, people try to set things up so that it's only favourable to them and you don't want your evaluations to be seen beneficial to the agency and not worth anything. When I was working in the Congress, even the really good studies that would come out of agencies would be distrusted by the Congress because they felt that they were just trying to make the agency look good. . . . You kind of get this balance between participation so that everyone gets a hand in it but with some standards so that the standard will keep the evaluation from appearing to be in some way capitulating to what the agency wants. This tension is always between telling the truth and hearing good news. (IP11)

Some participants also affirmed that the evaluation policy should set clear objectives for stakeholders' engagement in the evaluation process in order to strengthen the relevance and use of evaluation findings and explained how the evaluation policy can leverage participation in evaluation activities that will eventually increase their evaluation capacity. For instance,

the policy should inform stakeholders about evaluation and that it is being conducted and tell them what everyone should expect about these processes and involve people and require of them active input about the evaluation process and make that as a cycle so people develop the skills of conducting evaluation and participating in the process and providing feedback. Year by year you will see how the recommendations are being implemented and so people can actually see the value of the evaluation through active participation. When people participate in the evaluation, they can see the value of it. I have seen that myself and I think policy should support that. (IP3)

Policy must include aspects of participatory evaluation processes. I would say that training and workshops are good but they are not enough. The organization could train a group of stakeholders as evaluation consultants and provide them with in-depth training so they can actually take on the lead in conducting evaluation. (IP18)

According to another participant,

I've seen many cases where evaluations are done and come with my recommendations and management was required to act upon them. And now you have all of these recommendations and really what they're saying is it's all wishy-washy and they're just going through the

motions and it doesn't mean anything. It's all very formal. You really need to engage people in the process and make them really believe in it. Just using the policy to make evaluation mandatory isn't enough. (IP6)

One of the participants who worked with people with disabilities indicated that, if evaluation policy relies on a participatory approach and promotes the inclusiveness of all relevant actors in the evaluation process, more attention will be paid to marginalized, disadvantaged or less powerful stakeholders:

The government really should develop the capacity of organizations to conduct their evaluation in a participatory way because some of the stakeholders, like people with disabilities, should be included in the process to improve their outcomes. (IP14)

Other participants argued that stakeholders' engagement in evaluation clarifies the process and enables them to learn from their experience and to redesign activities and create new ones on the basis of their reflections on their experience. This increases the success of the role of the evaluation policy and builds capacity to respond to new and emerging demands. One participant put it this way:

I think the majority of people learn through experience. I think the idea is to do evaluation and then model how evaluation is done right, inclusive, formative, strategic, however you would define that. You want an evaluation that's not too huge, that's not too controversial on a topic and then actually engage the organization numbers in making sense of the data. So kind of de-mystify it in a practical experience around working with evaluation so they can see. (IP5)

One of the participants provided an example of where the evaluation policy was used as an educational and communication tool to facilitate engagement in the evaluation:

We developed an evaluation policy for the National Research Council. It was made by the federal evaluation policy because we had to meet those requirements, but we felt that it was not suited to our organizational needs and we really needed to spell out things like expectations that we had from program managers and how we felt they should be involved in evaluations. We want people to know right off the bat that this is what is expected of you when we evaluate your program. This is the process that we use. So it was fairly detailed and we used it mainly as an educational tool...and it outlines all of the different things that you can expect from us and that we will be expecting from you. So as a communicative tool it was extremely effective and it contributed to that early capacity development in those programs that had never evaluated before because they did not really know what evaluation was. So you are really starting from scratch where there is nothing and the evaluation policy gives you an early framework to start building on to develop the capacity of that particular program. (IP16)

Another participant also explained that

My experience is that people have been very willing to look at the kinds of things that evaluators do for their programs if the evaluators have explained to them what they are doing, how they are doing it and what good can come of it for them. Specifically, from the agency, how they can improve some problem that they've got? (IP2)

Finally, one of the participants concluded by stating that "Participation is crucial. It is everything. If you don't have the agreement of the people who are going to use it, who are going to endure the change, I don't think it's worth doing it."

Thus, evaluators can use the evaluation policy to take on a purposeful role in terms of stimulating evaluation capacity use by engaging stakeholders, by communicating with them throughout the evaluation process, and by identifying teachable moments when participants can learn and apply various evaluation skills.

Observations

An interesting observation, which I will further consider in the discussion chapter, is the strong focus placed by the interviewees on the capacity to use evaluation at the organizational and individual levels. This could be attributed to the fact that more than half of the interviewees are evaluation scholars and practitioners who have made considerable contributions to the advancement of the evaluation field over the last 30 years. Therefore, they are more concerned about promoting the value of evaluation and enhancing the capacity of evaluation use than about discussing the practical aspects of doing the evaluation. Although many of the moderating variables fall under the scope of the capacity to do evaluation at the organizational and the individual levels, many participants provided more details and examples when they discussed the ways in which the evaluation policy makes evaluations useful. Here is some protracted commentary to underline the point.

Just as well to give you some sense of where I'm coming from. I just think that evaluation needs to be fundamentally approached and rethought. The real value of evaluation is as a learning function. Evaluation should be the intelligence of the organization. At this point it's the coroner or semantic coroner. Evaluation is the guide once you're dead. They do an examination to find out what killed you. Evaluation is retrospective and it looks back. It's still stuck in the 20th century. I would say it's still stuck in the 80s. It looks back. It writes these huge reports that are boring and nobody reads. The world of development still uses the old criteria it's been using for the last 30 years and we no longer live in that kind of world. (IP2)

One of the major problems with programs that even spent billions of dollars and they evaluate their work is the lack of use of the outcomes of the programs. So I'm an evaluator and one of the big problems is how do you encourage use of evaluations and the outcomes of innovations in huge programs and this is a critical thing. Has anything been done in the program to encourage the use of evaluation, and for the most part there is not. There is no continuing use after the end of any program. Even though these programs go through rigorous evaluations it just doesn't happen and it is difficult to see how it will happen. Is it that organizations are so different and so dispirited and so context laden? (IP5)

We evaluators are not good at evaluating each other. Yes, we get an AEA or other great souvenir people in a nominal way because they're not so positive about the current evaluators, but you don't have a lot of humbling criticism of other people's evaluations. The closest that I know of was the battles against Fetterman and empowerment evaluation where Scriven and Donaldson and Patented had a series of major debates. (IP11)

Evaluation is a big industry. A lot of people including myself make money from evaluation. It is also a very large industry within governments and multilateral agencies around the world. It makes work, it makes work for people in departments and units and appoints people and so on and so forth. And quite often, for a long time, evaluation units were rare. Older staff people went to do nothing until they retired. There's still examples out there where people, they need another two years to get their pension, they need to ride it out and we need to think of any use from them in terms of productive use. So they're named the head of the evaluation unit so they can sit there and say all the right words and have evaluations stemmed and you're good to go. (IP15)

Summary

In this chapter, I described the macro-meso-micro ecological framework that I developed through the analysis of my findings in order to elucidate the multiple variables that moderate the role of evaluation policy in leveraging organizational capacity to do and use evaluation. The role of evaluation policy in building evaluation capacity to do and use evaluation is embedded within a socio-political context and is moderated by the type and content of the evaluation policy. Within these interconnected variables, organizational capacity to do and use evaluation is influenced in different ways. While organizational capacity to do evaluation is influenced by the presence of supporting structures, leadership, resources, and evaluation culture, the capacity to use evaluation is moderated by the conceptualization of the evaluation purpose within the organization and by the extent to which evaluation is integrated into the various decision-making processes. These variables further moderate the ways in which the evaluation policy influences capacity at the individual level. In terms of the capacity to do evaluation, the role played by evaluators and their interpersonal relations affect the potential influence of the evaluation policy. The

capacity to use evaluation is affected by the engagement of stakeholders in the policy-making and evaluation processes and, ultimately, by the learning function of the evaluation.

Most importantly, the framework describes the interconnected levels and variables that influence the role of evaluation policy in building organizational capacity to do and use evaluation. These levels and variables range from the socio-economic context of the country within which the organization is located through to the learning function and use of findings at the stakeholder level. These moderating variables provide a sense of how conducting evaluation and how the constructions of evaluation knowledge are implicated in different levels. The question of the relationship between evaluation policy and organizational capacity for evaluation becomes more tangible as the variables at the sociopolitical, organizational, and individual levels inter-weave throughout the evaluation process. In turning next to evaluation community members in Jordan and Turkey to test out some of these ideas, my understanding of the role of evaluation policy is rendered more complete.

Chapter 6: Findings from the Focus Group with MENA Evaluation Community Members

In this chapter, I present the findings obtained from the two focus groups that I conducted with evaluation community members and evaluation professionals located in the Middle East and North Africa (MENA) region, specifically in Jordan and Turkey. This study is intended to answer the third overarching research question as it provides a direct opportunity to explore contextual issues, whereas in Study 2 such issues were somewhat sporadic and incidental. Specifically, this study is focused on gaining an understanding in a practical sense of the contextual facilitators and barriers that influence the role of evaluation policy in enhancing the integration of evaluation into organizations. The purpose of the focus groups was to discuss with the members of the MENA evaluation community the broad themes that I had previously identified in Study 2 through my interviews with evaluation scholars and practitioners, as well as to explore the extent to which these themes resonate within this distinctive cultural context. My idea was to obtain feedback on the three-level ecological framework that was the product of my interview study (introduced in Chapter 5); I was interested in ascertaining whether the variables included in the framework made sense to these participants and whether they resonated with their experiences at the evaluation community level in relation to their evaluation practice and capacity-building processes within their respective contexts. Thus, it is important to note that these focus groups with the MENA evaluation community members were intended not merely as a way to triangulate the data collected in Study 2, but also as a way to incorporate the perspectives of the community members within these contexts into the understanding of the relationship between evaluation policy and ECB.

Major Findings

The findings reveal that there is a consensus about the relevance of the ecological framework across both of the focus groups and that there is agreement that this framework captures the variables that moderate evaluation capacity building within both countries. While the findings obtained from both focus groups correspond to all of the variables within the ecological framework, the discussions reflect greater consideration of the moderating variable in relation to the social, political, historical and economic context at the macro level

and in relation to its significant impact on the “capacity to do evaluation” at the meso level (organizations) and at the micro level (stakeholders) in both countries (See Figure 4, Chapter 5). Despite the differences between the Turkish and the Jordanian contexts, the focus-group participants felt strongly that the macro conditions associated with their countries’ political structures, social norms and values, historical characteristics and economic realities have significantly influenced and are continuing to influence evaluation practice and the various efforts that are being undertaken to develop evaluation capacity at the organizational and individual levels.

Most importantly, the participants in both focus groups stated very emphatically that they need an evaluation policy to help them build evaluation capacity and to bring people onto the same page with regard to valuing and practising evaluation. They expressed the belief that a well-thought-out evaluation policy would provide clarity around evaluation concepts, processes, principles, requirements, roles and responsibilities, use and accountability.

In what follows, I present the findings obtained from the respective focus groups as they relate to each level of the framework by discussing insights drawn from both focus groups in order to better explain the role and use of evaluation policy as well as its influence on evaluation practice and ECB activities.

Turkish Focus Group

Macro level: Social, political, historical, and economic context. In Turkey, the participants shared the view that, even though evaluation policy does not exist at the national level and even though evaluation is not a component of any strategic planning within government institutions, some kinds of evaluation-related activities started in the country in 2005 when the Turkish government began moving towards results-based management (RBM) and when organizations started to establish and integrate independent internal-audit systems. The Turkish participants commented that the development of fiscal and monetary policies and regulations tailored to the capital market as well as other changes in the latter were combined with more of a focus on audit practice. As a result, government organizations are required to perform independent internal audits in accordance with audit standards in

order to assess compliance with current regulations and demonstrate effective use of resources.

At the Turkish government level, the participants pointed out that the Turkish Court of Accounts carries out performance audits using a set of predetermined indicators in order to assess whether public resources have been used effectively. According to the participants, regardless of the designation of this type of audit as performance audit, the practice of the Turkish Court of Accounts is similar to evaluation practice. However, because the government is not particularly interested in evaluation, more restrictions have been imposed on this function. According to one of the participants,

The Court of Accounts does performance auditing. It was like an evaluation study. They ask about impact and about the effectiveness and efficiency just like an evaluation. But then the government said that you are asking these questions but this is not the responsibility of the Court of Accounts. You are only supposed to do auditing and not to evaluate impact. So what the government did is that they changed the name of the function from 'performance audit' and called it 'performance information audit' so that the Court of Accounts just provides information about the indicators and prepares activity reports.

It is worth mentioning that Wilkins and Boyle's (2011) review of the standards of performance auditing used by institutions in a number of countries suggests that there remains considerable controversy about whether performance auditing is embedded in the traditions of financial auditing or is based more on evaluation traditions.

The discussion with the participants indicated that the lack of interest in evaluation and the absence of demand for evaluative information for decision making at the government level form a barrier around the development of evaluation culture and capacity building in Turkey. Evaluation is seen as having no use in the Turkish government system. As one of the participants commented,

The government has no intention to set up an evaluation policy. They have no desire to be accountable and they have no desire to learn from such policy because they already know everything. On the other side, there is a history of auditing in the country, either internal auditing or auditing with the Court of Accounts, that is one of the most respected institutions in Turkey. They do auditing to all of the financial institutions and departments. So there is no need for evaluation.

Some participants attributed this approach to the prevailing social values and norms and how things are done in the Turkish culture. One of the participants explained that the lack of interest in evaluation as a learning and improvement mechanism stems from some common

cultural norms and perspectives that are based on trust—in particular, trust that the individuals who are in positions of power know their job and do not need evaluation to help them. For example,

In Turkey and in the Eastern culture, people are not interested in learning [from evaluation]. They know everything. They have to know everything because they are the top management. They are expected to know everything. If you are in this position it means you know everything.

Participants also commented on the economic condition of the country and affirmed that it is a critical moderating variable that influences demand for evaluation and drives the development of evaluation capacity at different levels. For example, countries that receive support from development and aid organizations are more exposed to evaluation activities that are conducted regularly to demonstrate impact and meet the donor requirements than countries that do not receive this kind of support. For example,

Some African countries have established their evaluation system because they receive donor money but Turkey is an emerging economy not receiving any funds. So nobody is asking for independent evaluation and there is no demand for it . . . when there is no demand it hinders the development of evaluation.

Although requirements stipulated by donor organizations might be seen as a demand-generating factor, it should be noted that donor requirements are not a substitute for local demand for and commitment to evaluation for capacity building.

One of the participants shared an important insight about the impact of the macro level on evaluation policy:

My point of view is that the policies of the country impact the evaluation. I mean the policymaking culture of the country itself. It impacts how to do evaluation, how to use evaluation, and how the leadership treats the evaluation. From my international development experience, the policymaking culture impacts how we formulate policies, how we implement policies, how we evaluate policies, and how this cycle is determined. Is it top down? Is it bottom up? This is very important for evaluation and the use of evaluation.

These general conditions at the macro level have proven to have impacts on evaluation capacity at the meso level.

Meso level: Organizational context. Participants in the Turkish focus group contextualized the influence of the macro-level moderating variables on organizational capacity for evaluation, especially on the capacity to do evaluation. They shared the view that leadership plays a critical role in promoting evaluation activities in Turkish organizations. For example, a number of the participants commented that, although evaluation is not mandatory in Turkey, managers could still ask for evaluation as a management practice or to inform decision making. As one participant explained,

If the manager wants to learn about the evaluation or the contribution of evaluation to management, then that manager will require evaluation. It depends on the manager. They can ask the internal auditor to do the evaluation for them and they can use the results of the reports, but these reports are not mandatory for the managers. They are just a contribution to management. They can use it for program and policy improvement.

On the other hand, it is common for managers to resist the mandatory internal audit, not to mention the evaluation. According to one participant,

Something is missing in the culture. Some of the top management don't even want audit so they will be, like, "you are evaluating me? My performance? And you're giving this information to the Court of Accounts! What kind of person are you? You are working in my institution." This is the kind of culture. So they would come up with their own strategies to prevent the auditors from making the audit or they don't recruit them or if they have them they give them something else to do. We have this hierarchy and the management is the problem.

When the participants were asked about the capacity to use evaluation, they shared the observation that there is a common misunderstanding of the meaning of evaluation and that evaluation is consistently confused with auditing. As I explained in Chapter 5, the conceptualization of the meaning of evaluation has a significant impact on how people perceive its usefulness. Such perceptions are mostly connected to the reasons why evaluation was introduced in the first place. One of the participants commented that

The biggest challenge for evaluation is that evaluation is seen as equal to audit. It is used for punishing people. Making evaluation popular is difficult because of this understanding.

Similarly, another participant commented on this common misconception that evaluation is similar to auditing, even among many university professors. This long-held misconception poses a challenge for evaluation and requires a cultural change as a prerequisite to being corrected. According to one of the participants,

Here [in Turkey] everybody thinks evaluation is equal to audit, everyone in the ministry of development and also in higher education and university professors. Maybe 50 years later they will see evaluation as a little bit different from audit.

Micro level: Evaluation stakeholders. At the micro level, participants commented on the interpersonal relations among various stakeholders and the ways in which they are shaped and influenced by variables at the macro and meso levels. It is evident that understanding power dynamics is an integral part of evaluation and capacity-building activities because this understanding can either enable or constrain change in the practice of evaluation. Indeed, an organization's power relations frame its conception of what evaluation actually means. Complex power dynamics may even hinder equal and genuine communication between different stakeholders. The following discussion among three focus-group participants provides a sense of the frustration that can accompany what some stakeholders consider to be a complex web of interpersonal relations affecting evaluation:

Participant 1: I think maybe in Canada there is a demand from the micro level that drives the need for evaluation and requires accountability and transparency. In our culture we don't have that. You can't even think of asking. I mean could you imagine asking the prime minister "How did you spend this money?"

Participant 2: This is the job of the opposition party to ask.

Participant 1: No, it is not the job of the opposition party. It is our job. I can't even ask my municipality. It is impossible to reach them and I'm a professor and I can't reach them. Even I vote for the party but I can't ask, "How did you spend the money?"

Participant 3: I think it is not the fear of asking. It is the culture that "we know everything." We also think the government knows everything, so why ask? I think it is trust or maybe because of the religion, you know, we have to obey.

Nevertheless, the participants still think that evaluators have a role in promoting evaluation culture and in educating people about the true value of evaluation and the benefits that it can bring to their organizations. They suggested that evaluators have to be creative and have to identify various partners with which to communicate, collaborate, and initiate dialogue in order to foster a culture of evaluation. One of the participants commented that

There are four actors eligible for evaluation in Turkey: government, Turkish NGOs, foreign NGOs, and UN organizations. NGOs in Turkey probably would be more experimental in developing evaluation culture than the government. I think maybe we need to look elsewhere other than the government to develop the evaluation culture.

As mentioned in Chapter 3, the focus-group participants are members of the newly established Turkish Monitoring and Evaluation Society. They are enthusiastic about evaluation and are values-driven in terms of promoting learning and bringing about positive change. One participant put it this way:

Evaluation minus learning equals audit. Then people understand it because they don't want the learning side. We are trying to create this culture here because we want the learning side.

Jordanian Focus Group

Macro level: Social, political, historical, and economic context. In Jordan, the members of the focus group shared the perspective that, even though there are monitoring and evaluation units in some government institutions and even though some evaluation activities are taking place in various organizations, there is no national-level evaluation policy to guide evaluation processes. Nevertheless, the participants noted that there is a government initiative at the national level named the *King Abdullah II Award for Excellence*. Launched in 2006 following the establishment of King Abdullah II Center for Excellence, this award is aimed at promoting performance improvement and awareness and at recognizing quality achievements nationwide. The award is the highest level of recognition for quality in Jordan and it is conferred every two years. The award's stated mission is

To become the national reference for quality and excellence among public, private, business associations, educational service providers and non-governmental institutions, measuring their ability and efficacy in serving their individual stakeholders and their contribution in raising the competitiveness of the Jordanian economy. (KACE, 2018)

Many participants commented that, despite the fact that the award acts as an incentive for organizations to improve the quality of their performance, the way in which this takes place does not inform practice. Thus, the award does not create a lasting commitment to continuous improvement. As one of the participants mentioned, for some organizations "the end goal is the award of excellence." Another participant also explained that

For the award you have to do this and that and if you don't do it, you are not excellent or not listed. They did not build the capacity within the organization and the staff themselves.

This means that, rather than being designed to develop sound practices and to institutionalize them within organizations, this award merely rewards the organizations and individuals that are selected.

Furthermore, in Jordan at the national level there is a strong presence of multiple aid agencies (e.g., UNDP, UNESCO, USAID) that are working on joint development programs with government organizations and that have their regional office in Amman, the country's capital. Participants shared the observation that the relationship with international development agencies and their involvement in development projects have influenced evaluation activities at the organizational level. As one of the participants mentioned,

I work in the Ministry of Land and International Population. It is one of the first government institutions to introduce the concept of evaluation to the government. This came from our work and relationships with the donors and international community. Learning about the evaluation function came out of our relationship with them.

Yet, the discussions with the participants revealed that there is no demand for evaluative information for strategic planning or decision-making processes at the government level. The following comments provide a sense of the kind of role for the government in promoting evaluation culture at the macro level that would have a significant impact on the organizational level:

Participant 1: The question is that is the government making evaluation important? Is it a priority? I don't think so.

Participant 2: I think the government should focus on evaluation. We have a problem of corruption in Jordan. We have corrupted institutions and if we don't have evaluation, we will continue falling down.

Participants also commented on the fundamental role of cultural norms and values and how the lack of an understanding culture influences evaluation and the perception thereof while acting as a major barrier to the development of capacity building at multiple levels, as one of the participants explained:

I think it is the culture of evaluation. We are Arab, and Arab people take things personally. People must accept and understand the value of evaluation and know that we are not evaluating them. We are evaluating the project. But people think you are not evaluating the project. You are evaluating 'my' project. There is a strong sense of ownership. Territorialism.

In sum, as is the case with the Turkish focus group, the discussion with the Jordanian focus group underscores the critical influence of the macro-level variables on evaluation and ECB activities. Thus, in order to understand the relationship between evaluation policy and organizational capacity for evaluation, it is important to capture the continuous interplay

among macro-level variables at the organization's external frontier that condition the role and performance of evaluation within the organization.

Meso level: Organizational context. Discussion with the Jordanian focus-group members confirmed the significant role of organizational leadership in enhancing evaluation. It should be noted that findings from all three of the studies of my research clearly demonstrate that leadership is essential to fostering evaluation capacity within an organization and that evaluation activities can be significantly encouraged and supported if there is informed and sensible demand for evaluative information from the organization's management. One of the participants provided an example to explain this point:

I stress leadership because evaluation was introduced in our organization by the management because they see its importance. They provided all the support and the budget needed to establish the evaluation department. We started with one person and now we are eleven. They are providing all the needed resources and currently none of our projects will reach the last stage without passing through the M&E department and integrating the M&E component and doing the budgeting and the costing for this component and ensure that there will be follow up. Yes it exists because the leadership believes in it and there is a budget for it.

This example provides an excellent illustration of the role of leadership in building organizational evaluation capacity. Clearly, when the management is interested in evaluation, it creates the demand, provides the needed support and makes use of evaluative information by integrating it into the decision-making process, as one of the participants emphasized:

Leadership is more important than the other variables because they can build the other things. I mean like the budget and staff training and capacity.

To further stress the role of leadership, another participant put it this way:

When I look at this figure, I would like to put leadership first and then leadership and then leadership. It is very important.

Leadership can support evaluation capacity and can inform demand for evaluative information within the organization by promulgating requirements that planning, budgeting, and reporting activities and documents be informed by evaluation. One of the participants explained this point by means of the following example:

The leadership is very important. In 2010 we had a prime minister who established a policy. You can even go to the website and see all of the governmental projects with deadlines and budget and percentage of achievement but after that prime minister is retired, the other one wasn't really interested in evaluation.

In fact, organizational leadership is a key to institutionalizing evaluation, as one of the participants commented:

The main problem is instituting the evaluation. If we have the leadership and the policy and if we have the demand for these evaluations we will be able to institute evaluation. We will use them.

Supportive leadership fosters not only the organizational capacity to do evaluation but also the capacity to use evaluation by integrating evaluation into the various decision-making processes. The participants' comments set out below show the level of frustration arising from the lack of evaluation use at the management level. All of the focus-group participants are of the view that the existence of an evaluation policy will inform management practice and will facilitate the integration of evaluation.

Participant 1: There is no appreciation of evaluators. There is no appreciation of the work they do and no decision making based on evaluation work. Nobody uses the results. We work and work and then the leadership changes or the minister changes. We evaluate for the sake of evaluation. There is a sustainability issue.

Participant 2: I think if we have a policy, the leaders will have no option. They will have to do evaluation and it will be integrated in the system and so we will have evaluations and reports and everything and it will be visible.

Other participants shared the opinion that, despite the existence of monitoring and evaluation units, evaluation is clearly not considered to be an important function within the organization:

You can see that organizations have a monitoring and evaluation unit but the reality is it is just a monitoring with no evaluation. It is not integrated through the organization like other functions.

This could be attributed to the fact that there is a lack of awareness of the meaning and value of evaluation and of the learning benefits that it could bring to the organization. As this participant explained,

The concept itself is an issue. No difference between evaluation and monitoring. We need to raise people's awareness about the meaning of evaluation. When we are talking about evaluation we need to focus on the learning and impact, not about following the procedures like do one, two, three.

Participants also believe that evaluation policy will enable the establishment of a system that facilitates the flow of information within the organization, increases coordination among

units and departments, and specifies roles and responsibilities for evaluation. According to the participants, the lack of clarity around these issues makes it very challenging for organizations to develop evaluation capacity:

Participant 1: If we start at the first step like ‘results’ even though evaluation is a higher level, but even if we want to start with the result we are struggling with the institutionalization in terms of how to collect data—who will collect data and for what purposes? Therefore we need a policy. A clear policy that has all of these components like who is responsible for what, what do we mean by evaluation and all of the other definitions.

Participant 2: We can do that for the [EvalJordan] association. I’m working on policy development for the health project now and we have a template. I think we can use it to develop a policy. I think we can start building our evaluation policy.

Clearly, the above-stated comment about creating a policy for their association is an indication of the value that the participants believe the evaluation policy would bring in terms of guiding the efforts for building organization evaluation capacity.

Micro level: Evaluation stakeholders. Discussion with the Jordanian focus-group participants provided insight into the impact of working with aid agencies on the development of capacity to do evaluation at the individual level. Many participants shared the view that their collaboration and relationships with international aid agencies helped to foster capacity to do evaluation by increasing evaluation knowledge and skills. One of the participants shared an example to explain how working with USAID cultivated evaluation capacity at the individual level:

USAID helps with enhancing the capacity of staff working on their projects and improving the engagement in evaluation. This has a good impact not only on the project and organization level but on the national level as well because we leveraged that here in EvalJordan and managed to build a relationship with the national evaluation department to deliver evaluation training sessions in our last conference and we will build on that because we need the sustainability. The manuals and procedures are coming from the USAID. They will be models for us to follow as we work on building our own capacity.

The above-stated example indicates that there is a strong focus on building the capacity to do evaluation because it is more readily under the control of evaluators, and thus it received more thought than the capacity to use evaluation, which depends on other variables like leadership. This point is evident in the following example:

We have a gap between the capacity to do evaluation and the capacity to use it. We have projects funded by USAID and they have an evaluation policy. How do we use it as implementers? Well, there is a gap and we shared it with the donor but they keep asking us to

use it for learning. They keep saying use it and use the results for learning but this brings us back to the importance of leadership. They don't use it. They put it on the shelf.

This participant also commented that there is a lack of understanding of the concept of learning and of what it means to learn from evaluation, especially when working with international development agencies:

We keep hearing about the concept of learning from the donor, but we also don't know what kind of learning. They want us to learn, and they are not only focusing on the results but the problem is the evaluation capacity of the implementer.

Many of the focus-group participants strongly expressed the view that the role of the evaluator is important for fostering evaluation capacity and for contributing to minimizing the gap between the capacity to do evaluation and the capacity to use it. Having evaluators with adequate skills and training to design and conduct evaluations, as well as to plan for and facilitate learning from the evaluation, is essential for promoting evaluation. The following comments underscore the importance of the evaluator's role.

Participant 1: The problem is the evaluators. The people who are working on evaluation they don't have knowledge or clear understanding of evaluation so when they write the reports the reports are junk. They are useless and with no value. Therefore the leaders don't take that report into consideration for decision making when they need to make a decision.

Participant 2: I think we have to decide where the problem started. Did it start with the rubbish that goes to the leaders or the leaders who don't support evaluation?

Participant 3: We have to build the capacity and keep the pace and provide the decision makers with good evaluation so they can see the value of it and then they will use it and support it.

It is clear that members of the focus group take responsibility for promoting evaluation practice and knowledge and realize that, as evaluators, they have an active role to play in fostering evaluation culture by providing credible and useful evaluation results. As members of EvalJordan, the participants also believe that their role is important for cultivating evaluation culture at the national level and so they act as advocates for the creation of an evaluation policy. Indeed,

We need to continue working on our own capacity. We need to raise awareness. We also need to identify stakeholders at the national level—people who need evaluation and will use evaluation. This will be a robust exercise. Then we need to use a change management approach and establish champions within ministries and within different levels of the

government and use these champions to take this effort forward with the vision of putting in place an evaluation policy.

Moving forward we have to identify the owner of this issue. Who will drive the development of evaluation? We need a national team to develop a system.

All of the focus group participants are keen on the idea of having an evaluation policy. They feel strongly that this is what they need in order to organize their efforts and move evaluation forward at multiple levels:

We have many plans and many guidelines. Things need to be filtered. We need one unified policy so everyone can go back to it if they have any questions.

Yes, we need one evaluation policy. A clear policy. A clear accountability. A clear action plan so we can move forward.

Finally, one of the participants concluded by stating “Let us encourage people that we are doing this for the benefit of everybody.” Thus, this confirms that evaluators can play an important role in communicating the benefits of evaluation and in strengthening its usefulness and relevance to various stakeholders.

Summary

In this chapter, I provide the findings from Study 3 of my research, which consisted of two focus groups that I conducted with evaluation community members in Turkey and Jordan. While the primary purpose of the focus groups was to better understand contextual influence and the perspective of the evaluation community members, and thus better understand the relationship between evaluation policy and organizational capacity for evaluation, these groups were also aimed at obtaining input on the ecological framework that I had developed based on my interviews with scholars and practitioners in Study 2 of my research (Chapter 5). Although there was considerable discussion about all of the variables included in the framework, the most substantive focus was on the moderating variables related to the capacity to do evaluation at the organizational and individual levels. Contextual variables touching on experience with evaluation, perceptions of its worth and fit, and evaluation demand, were shown to be quite influential. An interesting difference between Turkey and Jordan was Turkey’s lack of evaluation demand inasmuch as it is an emerging economy, not as heavily reliant on donor or international development funding. Jordan, on

the other hand, has a considerable history of working with external donors and therefore experience with evaluation.

Another interesting observation is that the participants in both focus groups stated that they believe that they need an evaluation policy. They also strongly expressed the view that an evaluation policy will support the development of organizational capacity for evaluation. All of the participants acknowledged that skilled, knowledgeable, and well-informed evaluators are essential for promoting the role of the evaluation policy and for facilitating capacity-building efforts and activities.

I conclude from this focus group study that the ecological framework arising from Study 2 has merit as a lens for examining the relationship between evaluation policy and ECB and that contextual variables are potentially powerful influencers on this relationship as either facilitators or, as was more to the point in the focus group data, barriers leveraging evaluation capacity from policy.

Chapter 7: Discussion and Conclusion

This chapter provides a summary of the key findings of the study and, through a thematic analysis, situates them within the broader literature on ECB and evaluation policy. In addition to the limitations of the study, further implications for practice and research are presented at the end of the chapter.

Summary of Findings

The three interconnected studies have each provided a different lens through which to explore the multiple dimensions of evaluation policy in relation to evaluation practice and organizational capacity to conduct and use evaluation, and through which to further explore what factors come into play as either facilitators or barriers to drawing evaluation capacity from evaluation policy within different contexts.

In the first study (Chapter 4, Research Question 1), my focus was 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 having a positive impact on the evaluation function and on organizational capacity for evaluation. One key finding of this study is the refinement of Trochim's (2009) evaluation policy framework, which includes eight evaluation policy categories. It should be noted that Trochim's evaluation policy framework was generated from a theoretical standpoint, rather than on an empirical one. Based on this study, I was able to identify 10 broad categories that capture the goals, processes, strategies, criteria and activities of evaluation policy, of which 8 correspond with Trochim's categories (with some adjustment and integration) and of which two are additional categories that stood out during the analysis as being employed in a number of policies.

These categories capture the dynamic range and multiple manifestations of the evaluation processes within various organizations and are supposed to have a direct influence on organizations' capacity to do and use evaluation. Set out hereunder is a summary of the categories.

- 1- **Evaluation Goal** refers to the purpose of the organization's evaluation function. Ensuring accountability, promoting learning and developing/supporting an evaluative culture are the three major goals of evaluation that emerged in association with this evaluation policy category.
- 2- **Evaluation Definition** refers to the definitions included to clarify the meaning of evaluation and to distinguish between evaluation and other closely related concepts and functions within organizations, such as auditing, inspection, and measurement.
- 3- **Evaluation Role** refers to the evaluation responsibilities of various organization stakeholders. The data indicate that clearly defined roles and responsibilities are important for carrying out the evaluation function so that every individual who is involved in the evaluation is clear about his or her responsibilities.
- 4- **Evaluation Management** refers to the management of the evaluation function within the organization, including evaluation budgeting and resource allocation, evaluation-related human resources management and evaluation timing.
- 5- **Evaluation Norms and Standards** refers to the evaluation-related norms and standards that are accepted as correct within an organization. It aims to foster an organizational culture in which members are aware of and agree upon the value of the evaluation function and its role within the organization.
- 6- **Evaluation Planning and Processes** consists of evaluation design and planning, evaluation methods, stakeholders' participation and evaluation with partners. It focuses on the processes and procedures of planning and doing evaluations.
- 7- **Evaluation Use** focuses on the reporting, dissemination, and use of evaluation results. The majority of policies indicate that the identification of evaluation use starts with the planning stage of the evaluation and requires that every evaluation plan identify the users, involve them in all of the evaluation processes, and indicate the ways in which the results are to be communicated, used and disseminated.
- 8- **Meta-evaluation** focuses on the review and evaluation of evaluations (and in some cases evaluators) as part of the organization's evaluation management and quality control system.

9- **ECB** focuses on the range of efforts that an organization may undertake in order to integrate evaluation practices and use them at all levels of the organization and thus strengthen its operations and improve its activities.

10- **Competencies and Ethics** focuses on the qualifications, skills, expertise, and ethics that organization managers and staff, especially evaluators, require in order to fulfil their roles and responsibilities for the evaluation and thus ensure the credibility and quality of the evaluation processes and products.

Another key outcome of this study is a revised definition of evaluation policy that is more comprehensive and that captures the meaning of the evaluation-policy construct as it is manifested in the extensive sample of evaluation policies included in this study. Based on my analysis, the definition of evaluation policy is as follows:

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.

The definition that was provided by Trochim (2009) focuses only on *doing* evaluation. This study highlights the fact that evaluation policies are concerned with much more than “rules and principles” and they guide decisions and actions that start before and extend beyond the stage of “doing evaluation” as is suggested by the revised definition.

In the second study of my research (Chapter 5; Research Questions 1, 2, & 3), my focus was on the further exploration of the relationship between evaluation policy and organizational evaluation capacity and the factors that affect evaluation policy's role in the integration of evaluation within the organization culture. My findings show that, in and of itself, evaluation policy has little relationship with an organization's capacity to implement evaluation or to use it, though this is not always the case. The majority of the interview participants see the role of evaluation policy either as symbolic or as an organizational means of enforcing control. The findings of the study also reveal that certain variables moderate the influence of evaluation policy on ECB: these moderating variables operate at three different levels (see Figure 4 in Chapter 5). The three-level ecological framework that emerged from

this study is intended to depict the key variables that moderate the relationship between evaluation policy and organizational capacity for evaluation.

The macro-level factors include the socio-political, historical and economic context of the organization and the characteristics of often externally imposed evaluation policies that enhance or impede these policies' influence on the ECB of organizations and individuals in terms of their capacity to do evaluation and to use it. The presence of a supportive system, the availability of resources, the commitment of the leadership, the presence of an evaluative culture that accepts and values evaluation, and the plan and processes for implementing the evaluation policy are all critical variables that moderate evaluation policy's influence on the *capacity to do* evaluation at the organizational level (meso level), which in turn influences the role of the evaluator and the interpersonal relations among stakeholders at the individual level (micro level).

The integration of evaluation into decision-making processes and the conceptualization of the purpose of evaluation are key variables that moderate the role of evaluation policy in influencing the *capacity to use* evaluation at the organizational level. These variables in turn influence stakeholders' engagement in the evaluation process and their learning from evaluation activities, these being the two variables that were identified as moderating the influence of policy on the capacity to use evaluation at the individual level.

In the third study of my research (Chapter 6, Research Question 3), the primary focus was on exploring the evaluation community members' and the evaluation professionals' perspectives related to the three-level ecological framework identified in the previous study. One of the chief findings about evaluation policies in this study is that there is broad agreement among the members of the Turkish and Jordanian focus groups that the framework is relevant and that it captures the variables that moderate evaluation capacity building within both countries. Finally, the members of both focus groups affirmed that they have an acute need for evaluation policy to guide their efforts in building evaluation capacity in their respective contexts. They expressed the belief that well-thought-out evaluation policies (national, organizational) would provide clarity around evaluation concepts, processes, principles, requirements, roles and responsibilities, accountability and, ultimately, use and better situate evaluation culturally.

Relevant Themes Emerging from the Literature

Many of the themes identified in the literature review presented in Chapter 2 can be informed by the findings of this research. In this section, I revisit key areas from the original literature review in order to highlight relevant findings and establish linkages between some of the key concepts.

Evaluation Context

In real estate, it's location, location, location.
The evaluation counterpart is context, context, context.
Marvin C. Alkin (Vo & Christie, 2015)

This research study characterizes the relationship between evaluation policy and organizational evaluation capacity as a cyclical process that involves knowledge generation and is aimed at improving evaluation capacity. Most importantly, my initial conceptual framework (Figure 1, Chapter 2) acknowledges the fact that ECB is context dependent and that contextual factors are likely to influence all aspects of the process, either as facilitators or as barriers.

Over the course of the three successive studies of my research, the indispensable role of context and its influence on the relationship between evaluation policy and organizational capacity for evaluation became more and more evident. The findings of this research emphasize the fact that the socio-political, cultural and economic context of an organization is very critical and that it influences all aspects of the evaluation—to the extent that it is, in fact, the most salient feature of the ECB process. This finding is consistent with the recent work of scholars who call attention to the importance of context for evaluation practice (e.g., Chouinard & Milley, 2016; Vo, 2013; Vo & Christie, 2015).

Although only a limited number of scholarly studies shed empirical light on the influence of context on evaluation practice and evaluation use, many scholars have argued that the acquisition of an understanding of the contextual factors pertaining to the socio-political, economic and historical context of an organization can generate insights that are important for explaining how interaction between hierarchies, systems, structures and people influences evaluation processes and evaluation use within organizations (Chelimsky, 2012; Vo, 2013; Vo & Christie, 2015). For example, in one of the recent studies about context, Chouinard and Milley (2016) conducted a thematic comparative analysis of two reviews of

the empirical literature on participatory evaluation in order to explore the multidimensional significance of context for the practice of CAE. The authors identified five contextual dominations that influence the collaborative evaluation process, namely the historical dimension, the epistemic dimension (how knowledge is constructed), the political dimension, the cultural dimension, and the organizational dimension. The authors ultimately argued that consideration of the dynamics of the contextual dimensions “allow for a refashioning of the contours and possibilities of evaluation research” (p. 8).

In relation to the context of international development, Raimondo (2018) discusses several gaps in the literature pertaining to the use and influence of evaluation and emphasizes the fact that the existing evaluation theories in these areas are based on models of rational organizations that ignore contextual factors like institutional norms, routines and belief systems. The author further argues that important insights from international organization theory allow for better understanding of complex conjunctions of context, including those of material, cultural, internal and external factors affecting evaluation on the one hand, and processes of change at the organizational and environmental levels on the other hand.

This influence of context is salient in my research and is illustrated through the emergent ecological framework presented in Chapter 5 (Figure 4). The majority of the participants in Study 2 and Study 3 reflected on how governments, social policies, legislation, national economic conditions and culture are all contextual factors that moderate both the role of evaluation policy in building organizational capacity for evaluation and the ways in which people perceive evaluation in their respective countries. For example, being from the Middle East, I can argue that for most people in this region, the very concept of evaluation is problematic. The fundamental reason is simply that the idea of evaluation as it is perceived and understood in the West does not align very well with the Middle Eastern culture where preserving dignity and ‘save face’ are more significant and have a higher value than evaluating and judging one’s work.

As was evident in the findings from Study 3, Middle Eastern people in their respective societies are expected to show personal integrity in order to be socially accepted. They manifest honour and integrity by projecting a public impression of dignity derived from an ostensible lack of guilt. Even if the facts and conditions point to the contrary, the social image of non-guilt must be preserved if they are to maintain the socially expected face.

Dignity and respect are granted to those who show themselves flawless; in general, Middle Eastern societies in general accord no respect to people whose faults or errors come to public knowledge. Within this culture, blame, fault or error is likely to cause a fall from social grace and a loss of dignity or face; thus, Middle Eastern people generally feel revulsion and bitterness towards anything that might tend to compromise them in this way, including evaluation. For this reason, within this cultural context it is critical for evaluation policies to frame evaluation in a way that emphasizes its learning benefits rather than mainly focusing on accountability.

In fact, failure to recognize and fully understand the various contextual factors that support or impede capacity-building efforts can result in an evaluation policy that inadvertently undermines stakeholders' ownership of the development process by creating a vicious cycle of disempowerment and a vicious circle of demotivation within organizations (Theisohn & Courtnadge, 2005). For example, the use and abuse of power constitute one of the aspects of context that arguably is the most important to evaluation policy but also is the most seldom analyzed in the evaluation literature. Evaluation policies that are solely aimed at enhancing accountability and legitimacy may actually exacerbate entrenched power-differential issues, especially in organizations operating in contexts of monarchical, non-democratic or dominant-party governments. In such contexts, as was proven by the Study 3 findings to be the case in Jordan and Turkey, there is a long history of hierarchy as the natural order of things; high power distance is one of the principal features of the relationship between officials and peasants, rulers and ruled, and elites and masses. Therefore, a lack of understanding of this contextual issue may result in evaluation policies that support the preservation of the existing power imbalance, and this can have a negative impact both on evaluation and on ECB processes. Thus, the emergent nature of the capacity-building process within these contexts highlights the need for evaluation policies that address complex systems and that encourage flexibility, adaptability and innovation.

For evaluation policy to have a positive impact and, in particular, a beneficial influence on evaluation capacity in any context but particularly non-western, development ones, thorough contextual analysis should be recognized as a precondition for the provision of effective support to organizational capacity development processes. Such analysis can help with developing and disseminating understanding, first, of where prevailing gaps exist in the

organizational capacity to do evaluation, including in human-resources, organizational and administrative systems, and, second, of why these gaps exist in relation to socio-political, economic and cultural factors, as well as specific statutory and regulatory systems, that constrain or enable capacity-building efforts. This means that evaluation policies that are based on the premise that capacity building is basically a matter of ‘replicating best practices’ regardless of organizational context are not effective and are even problematic, especially in organizations that lack the capacity to support capacity development activities.

Thus, the incorporation of context into efforts to understand the role of evaluation policy in ECB provides a more holistic picture of the components of the process and has the potential to generate insights into ways of leveraging change using evaluation policy that might otherwise be overlooked. This approach also places the ongoing study of evaluation policy on a good footing relative to other contemporary domains of inquiry in the field of evaluation.

Evaluation Use

Evaluation utilization is important for building organizational capacity for evaluation because it is through the use of evaluation findings and through participation in evaluation processes and activities that the likelihood of organizational development increases (Patton, 1999, 2007, 2008). There is evidence indicating that many of the organizations that recently revised their evaluation policies did so in order to focus on the aspects that would increase various types of evaluation utilization including the conceptual and instrumental use of findings, and process use.

For example, some evaluation policies that encourage the conceptual use of findings highlight the importance of evaluation users’ involvement in increasing awareness and understanding of program issues and in addressing these issues; this in turn increases the likelihood that informed decisions will be made and that beneficial changes will occur. Furthermore, many of the examples drawn from the policies demonstrate that there is a clear growing emphasis on process use as another mode of learning benefit. Specifically, some policies outline the benefits and consequences of stakeholder involvement in the evaluation process in terms of increased learning and increased ownership (Alkin & Taut, 2003); this can have a great impact in terms of organizational ECB. Through process use, organization

members “learn to think evaluatively so as to build evaluation into the organizational culture and engage in evaluation as part of ongoing organizational development” (Patton, 1999, p. 94). At the same time, an emphasis on instrumental use (the use of evaluation findings) is noted in policies that provide details on the implementation of evaluation findings and the application of recommendations. In this sense, organization members transform what they have learned into action, reflected most often in discrete organizational decisions.

I also observed that several evaluation policies include elements pertaining to participatory or collaborative approaches to evaluation (in which stakeholders are actively involved in the planning, design and implementation of the evaluation), indicating that evaluation use starts with the planning stage of the evaluation, and requiring that every evaluation plan identify the users, involve them in the evaluation processes, and indicate how the results are to be communicated, used and disseminated. In fact, a growing recognition is reflected in the evaluation policies that top-down evaluation designs that do not take into account the information needs of stakeholders are less useful and are far less likely to succeed. This 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, 2013, 2015; Shulha et al., 2016).

Participants in Study 2 and Study 3 shared the view that the role of evaluation policy in fostering the type of evaluation use that increases learning and development at the organizational level (the meso level) is moderated by two interconnected variables: (1) managers who believe that evaluation is an important and meaningful practice for informing decision-making processes and (2) the purpose of the evaluation function within the organization. Based on the findings of this research, these two variables warrant more attention and are discussed in detail in the following sections.

Evaluation Purpose: Accountability and Learning

The analysis of the extensive sample of evaluation policies in the first study of this research demonstrates that evaluation is seen as a mechanism for the promotion of both accountability and learning. On the one hand, evaluation scholars and practitioners in the second study clearly insist that evaluation policies are mostly symbolic and accountability-

driven and that the requirements of evaluation policies add to the existing tension between the accountability-oriented needs of funders and overseers and the learning needs of other evaluation stakeholders such as managers and organization staff. In fact, this tension between accountability and learning as reasons for and results of evaluation has been a subject of interest for a number of scholars in the recent years because of its significant impact on evaluation practice and on the advancement of the evaluation field (Chouinard, 2013; Feinstein, 2012; Guijt, 2010).

On the other hand, the evaluation community members and evaluation professionals who participated in the third study had strikingly divergent opinions and affirmed that they need evaluation policies to guide their ECB efforts and support the development of organizational capacity for evaluation in ways that foster learning. The issue raised in this research pertains to the location of the tipping point between developing evaluation policies that encourage learning and foster organizational evaluation capacity on one side and formalizing evaluation policies that lean heavily towards accountability on the other. Part of the answer to this question lies within the very same purpose of evaluation—that is, accountability and learning. While today both accountability and learning are considered important motives for evaluation and are indeed cited as evaluation goals in most evaluation policies (Kogen, 2018), the challenge with these two concepts is that they are not necessarily clear and that their meaning, especially in the evaluation policies, is largely assumed rather than being explained.

I observed that the majority of the evaluation policies included in this study that cited learning as a goal for evaluation either do not include a clear definition of learning or refer to learning as a by-product of accountability that is mostly used to enforce even more accountability. Indeed, the general impression that emerges from the review of evaluation policies is that learning simply means ascertaining whether a project succeeded, which is specifically the aim of accountability. While accountability is concerned with determining whether a project has succeeded or failed (Lehtonen, 2005), the kind of learning that is referenced in the evaluation policies suggests that, by determining the success or failure of a program or project, the organization will find out which project or program to continue and which project or program to discontinue. Thus, this means that learning as a goal of evaluation is included in evaluation policies as an outcome of accountability and is mainly

concerned with assessing project success. For example, this is clearly stated in the most recent IFAD evaluation policy: “Accountability is considered as a necessary first step in the learning process” (IFAD, 2011, p. 9). This means that learning as an evaluation policy goal is not about what the organizations hope to learn from evaluation or about what program changes are warranted, which types of evaluation activities would be most useful, and so forth; this may explain why the majority of the interview participants in the second study were skeptical about the role of evaluation policy in building organizational capacity to use evaluation.

Moreover, one of the concerns that the participants shared is that this continuous emphasis on accountability can have considerable adverse effects on the evaluation function because it not only undermines the learning value of evaluation but also leads to symbolic use. This is because organizations are only producing *positive* evaluations that satisfy the funders’ needs in order to secure more funding, an approach that increases the incentive for “strategic misrepresentation” of evaluation findings (Lehtonen, 2014). As such, evaluation is a time-consuming task that is considered to be irrelevant for internal use within the organization (Ebrahim, 2005).

This ongoing tension between accountability on the one hand and the misunderstanding of the true meaning of learning from evaluation on the other hand poses challenges for the evaluation field in terms of finding ways to reconcile these two sides, especially given the ever-increasing demand for accountability (Armytage, 2011; Lennie & Tacchi, 2014; Regeer et al., 2016). In fact, it is for this reason that many evaluation theorists assert that “accountability-based” and “learning-based” evaluations are different from one other (Armytage, 2011; Lennie & Tacchi, 2014). Yet, if the focus in evaluation is consistently and increasingly on accountability and the question of whether or not a project has succeeded, it is important to ask when and under what circumstances organizations use “learning-based” evaluations.

Thus, a reconsideration of the definition and conceptualization of learning and accountability as they appear in evaluation policies is needed. Regeer et al. (2016) explore ways of reconciling accountability and learning by examining the differences in the needs and expectations of the various stakeholders who are involved in evaluation. The authors identify three types of accountability—upward, downward and internal—that evaluations are

supposed to address. Upward, or financial, accountability is concerned with accountability to funders (i.e., donors, investors, governments) and, in relation to this form of accountability, evaluation is used to ensure desired results. For upward accountability, the emphasis is on goal-oriented evaluation, and learning may occur through consideration of the evaluation outcomes. Downward accountability is to the intended beneficiaries of a program, and it entails involving relevant stakeholders in a participatory process at different stages of the evaluation, specifically in the assessment of the impact of the program and its desirability. Finally, internal accountability is related to the processes of organizational learning and change. This type of accountability emphasizes the importance of evaluation in terms of contributing to internal learning processes and decision-making within the organization; therefore, the focus is on conducting evaluations in such a way that the results can be used in practice (utilization-focused evaluation) (Patton, 2008) and are aligned with the learning needs of the organization members (responsive evaluation) (Stake & Abma, 2005) in order to foster learning and development throughout the evaluation process.

This conceptualization of accountability underscores the importance of rethinking the role of evaluation policies as mechanisms for promoting learning and for building capacity to use evaluation. For evaluation to be useful, it must be aimed at the right users. Given the vagueness and ambiguity surrounding the meaning of accountability and learning referenced in the vast majority of evaluation policies, the tensions between accountability and learning will continue to exist and these will serve to render the evaluation process superficial and in some cases even meaningless for capacity development. Clearly, while the participants in Study 2 expressed concerns about evaluation policy being used as a tool for reinforcing more *upward* accountability, as is the case in their respective contexts (i.e., the United States, Canada, Europe), the Jordanian and Turkish participants in Study 3 indicated that they are keen to have an evaluation policy to help them build organizational capacity and establish *internal* accountability for development. This finding is critical for the role of evaluation policy in building organizational capacity; evaluation policies have to clarify the purposes of evaluation and to address this tension. Otherwise, confusion will persist, and the weight of accountability demands will continue to thwart the potential for learning through evaluation to leverage organizational reform.

Organization Leaders' Evaluation Capacity

Five of the 10 evaluation policy categories identified in Study 1 speak directly to the important role of managers in fostering evaluation capacity. These are: (1) evaluation role, (2) evaluation management, (3) evaluation use, (4) ECB, and (5) competencies and ethics. 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.

The participants in Study 2 and Study 3 further emphasized the indispensable role of leadership in fostering the capacity to do and use evaluation. The findings show that for the policy to leverage organizational capacity for evaluation, the leadership must be seen as supporting and promoting evaluation. Leaders can do this by: (1) understanding how strategy and evaluation are interconnected (or should be interconnected); (2) providing adequate resources for evaluation; (3) being active consumers of evaluation information; and (4) using evaluation as a means for ongoing organizational learning (Preskill, 2014). This includes placing a high value on evaluation by consistently seeking evaluative information for planning, implementation, and decision-making processes—in other words, they have to walk the talk (Norman, 2002).

The findings also suggest that, within the ecological framework, leadership is a variable that moderates the role of evaluation policy in enhancing organizational capacity for evaluation and that forms a link between the capacity to do evaluation and the capacity to use it at the organizational level. In other words, organizational leadership is identified as a variable that is critical for supporting or leveraging the capacity to effectively plan and implement evaluation, on one hand, and that is responsible for integrating evaluation into decision-making processes under the capacity to use evaluation, on the other hand. This is because leaders (especially senior organizational leaders) have the power and the responsibility to support the organizational capacity for evaluation.

It is entirely evident that the contextual factors, including the socio-political, cultural, historical and economic characteristics of a country, have a direct influence on the leadership style of organizations within the context. After taking into account the findings of Study 2 and 3 and after reflecting on the role and style of leadership in developing countries, I identified several interesting characteristics that are worth mentioning in relation to ECB.

Within Middle Eastern societies in general, a leader is perceived as someone who “knows it all” because of his position, expertise and achievements. A leader has all the power to make decisions; in fact, it is likely that people would perceive leaders as being weak or even incompetent if they overly engage with other organization members in the decision-making process. Instead, leaders are expected to maintain a consultative approach where they consult other organizations’ members informally but make their own decisions unilaterally to show their power. Such power is associated with their status and position as leaders within the organization. Leaders are also expected to work on fostering their interpersonal relations with employees and to become involved in different aspects of their lives to gain trust and loyalty at a personal level. This kind of interaction between leaders and subordinate organization members is referred to as paternalistic leadership (Jackson, 2016) because it resembles the parent-child relationship. The paternalistic relationship is hierarchical as the leader assumes the role of a father who is responsible for staff and who is expected to exercise relational power that is more authoritarian and disciplinarian towards them, whereas the staff voluntarily are expected to submit to the leader (Mansur, Sobral & Goldszmidt, 2017).

It should be noted that, in the Middle East and in most developing country contexts, leaders are highly status conscious. They want (and mostly can) remain in a position of power to maintain their status in the society. Therefore, they are most likely to resist change in order to ensure that they do not lose power or relinquish authority. It is interesting to note that such leadership style is focused only on the work the leader expects and never on people. As such, it is more likely to hinder capacity building efforts. Given this instrumental role of leaders within organizations in these contexts, it follows that Western models and styles of leadership that are based on and are developed for democratic societies are inapplicable in such a highly hierarchical system.

The presence of supportive and committed leadership to facilitate and manage the implementation of evaluation capacity building efforts and to internalize norms and ownership of the process to ensure sustainability is the main element in the black box of what drives capacity building within these organizations, especially in developing-country contexts. However, it should be noted that the factors that generate this leadership themselves constitute yet another black box; the individual evaluation capacity of organizational leaders

is clearly the most salient feature of the process. Yet, as is demonstrated by the findings of Labin et al. (2012) with respect to ECB, “Leadership was the least frequently targeted organizational factor and the least frequently reported organizational outcome” (p. 321).

As my research findings confirm, to be able to support and promote meaningful evaluation within the organization, leaders must understand the value of evaluation as a management and decision making mechanism so they can develop the skill to think evaluatively and learn how to balance accountability and learning. In fact, organizational leaders have a great deal to learn. At the same time, as Preskill (2014) notes, “I do not think we have paid enough attention to the role senior leaders play in organizations, and how they influence, shape, and sustain an evaluation and learning culture” (Preskill, 2014, p. 117). My research underscores the need to delve more deeply into leadership as a moderating variable and to develop considerably more than is presently the case, how leadership might play out in cross-cultural contexts.

Evaluation Policy as a Capacity Building System

Donella Meadows begins her book *Thinking in Systems* (2008) by stating:

A system is not just any old collection of things. A system is an interconnected set of elements that is coherently organized in a way that achieves something... [A] system must consist of three kinds of things: elements, interconnections, and a function or purpose. (p. 11)

Developing evaluation capacity has to be a shared concern of the entire organization. Organizational leaders must be convinced of the need for evaluation as a learning mechanism to produce knowledge that is useful for their management role and for the development of the organization. For this to happen, evaluation requires a policy that enforces *internal accountability* and acts like an organizational system that is embedded in the organization’s arrangements. Organizational evaluation capacity often is most fully developed in organizations that have conceived of it as an integral part of the organization system and as *the way of doing things* (Labin et al., 2012; Mayne, 2010).

The findings of my research reveal that, if the policy incorporates and sends the message that the impetus for evaluation is entirely external and that the purpose of evaluation is mainly to meet upward accountability requirements, its role in building evaluation capacity is likely to be weak. The research demonstrated clearly that espoused evaluation policies

appear to depart quite considerably with what is observed in practice (i.e., policy-in-use). For evaluation to be taken seriously, there has to be a wide recognition within the organization that evaluation is about much more than satisfying an obligation to someone else. Evaluation policy without a clear understanding of the purpose of evaluation, committed management, resources, and supporting infrastructure and organizational culture will not be effective. In this kind of scenario, there will be token compliance rather than wholehearted adoption of evaluation practices. On the other hand, for organizations that do not have evaluation policies to guide and manage their practices, their efforts to develop organizational evaluation capacity are more likely to be fragmented and scattered.

This theme is interconnected with accountability versus learning considerations, as well as those associated with organizational leadership. If evaluation policy is to gain purchase as legitimate leverage for organizational evaluation capacity building, the learning function of evaluation must necessarily be emphasised. The ECB research of Cousins and Bourgeois (2014) generated a working hypothesis suggesting that ‘data use leads to data valuing’. The essential idea is that through a combination of pressure and support, if organization members are exposed to the successful use of evaluation, they will be more likely to embrace it as a potentially powerful strategy through which to leverage change. Evaluation policies that favour mechanistic accountability outcomes and concerns are not likely to be particularly meaningful to organization members. Yet their exposure to overt evaluation learning through, for example, organizational or program changes arising from evaluation findings or changes in thinking arising from evaluation processes is much more likely to enable members to see firsthand the benefits of evaluation.

Organizational leaders, who show a commitment to the learning value of evaluation through overt actions and responses to evaluation findings and processes, are well positioned to reduce the gap between espoused policy and policy-in-use. As covered in the foregoing literature review, policy development goes through several stages from inception to completion, such as agenda setting, formulation, legitimation, implementation and evaluation. (e.g., Anderson, 2011; Knoepfel et al., 2011; Smith & Larimer, 2009). As part of a commitment to organizational policy for evaluation capacity building, there is a role for organization leaders to not only help shape policy in terms of its emphasis on evaluation’s learning function but also to ensure that such emphasis is consistently evident in ensuing

stages, particularly, when the focus is on policy implementation and evaluation. Of course, as my findings demonstrate, a commitment to the learning function of evaluation needs to be understood within the complexities of context, particularly in cross-cultural contexts. As discussed above, the dynamics of organizational leadership in development contexts are substantially distinct from those of Western cultures where challenges to the foregoing sentiments about learning from evaluation may be relatively more problematic.

Limitations of the Study

The context within which this research was undertaken entails a number of limitations on the interpretation of its findings. The first limitation is associated with the fact that the majority of the policies included in the review and analysis in Study 1 came from international development organizations and aid agencies (mostly UN organizations). Given the context, size, scale of operation and policy requirements of organizations of this type, it follows that the examined evaluation policies are probably not independent of one another and that there probably has been some degree of diffusion – with an organization basing its new or revised evaluation policy in part on the existing policies of other organizations. Thus, the degree of variance has been somewhat constrained and the relative counts of the various evaluation policy categories may have been affected by whether or not a given category was included in an early policy or in a subsequently widely shared policy.

Furthermore, although the sampled evaluation policies span the globe (with organizations located in North America, Europe, Africa, Australasia, East Asia, and Latin America and the Caribbean), the likely omission of some evaluation policy components – specifically those that were not found in the organizations studied but that nonetheless exist in other organizations – could minimize the applicability of the evaluation policy categories to other types of organization (e.g., non-profits, schools).

The second limitation is related to the interview participants in Study 2. The individual interview format selected for this study aimed to explore, from a general perspective, the participants' overall experience and perception of evaluation policy, as opposed to focusing on the specific context in which each participant practiced evaluation. This approach, however, yielded results that did not provide detailed identification of the effects of the practice context on the views reported by the participants. It also did not

acknowledge the type, nature and scope of the evaluation policies to which the participants were referring. This limitation was mitigated by including verbatim quotations of comments made by some of the participants about their organization's type in order to obtain some level of representation of multiple organizational contexts.

Another limitation related to the participants in Study 2 and Study 3 is the fact that the participants in Study 2 were evaluation scholars and practitioners who have made significant contributions to the advancement of the evaluation field over the last 30 years – either in the ECB or in the evaluation-policy streams of inquiry – and the fact that the participants in Study 3 were evaluation community members who are interested and invested in ECB. Even though the participants in these two studies provided valuable insights into the role of evaluation policy in building organizational capacity for evaluation, their views were based on their own knowledge of evaluation. Therefore, while the ecological framework effectively represents the points of view of evaluators, the study could have benefited from a more deliberate inclusion of evaluation users, such as organization managers and senior decision makers, to ensure the presentation of multiple points of view within the ecological framework. Nevertheless, I can say with confidence that the findings of this study do in fact represent multiple points of view and, in particular, reflect the fact that the participants come from many different regions and backgrounds.

The third limitation is associated with the interpretation of the findings in Study 2. When I developed the initial conceptual framework (Figure 1) for my research proposal, I did not identify the relevance of the social ecological perspective (Bronfenbrenner, 1979, 2005) as a lens through which to understand the relationship between evaluation policy and organizational capacity for evaluation; this omission was made notwithstanding the fact that much has been known about this perspective for approximately 40 years. Rather, the ecological concept emerged from my analysis of the interview data and proved to be relevant in making sense of the findings. This limitation was mitigated by validating the ecological framework with Study 3 participants in two different countries. It should be noted that the research was designed to allow understanding to develop and unfold – hence the emergence of the ecological perspective during the data-analysis stage of Study 2.

The final limitation is related to my efforts to ensure the trustworthiness of the data. I took deliberate steps to enhance the trustworthiness and quality of my research data,

including explicitly detailing policy content in an appended table, developing and revising coding systems for the policy document analyses well as for the interview analyses, piloting interview and focus-group guides, generating verbatim transcriptions (as opposed to summaries) of interviews and focus groups, and having separate investigators conduct the focus groups while I closely attended to the participants' contributions and reactions to various suggestions and points of view. Yet, as an individual researcher with some limitations on access to resources, I did not engage with some of the common strategies used to enhance trustworthiness, such as assessing inter-rater agreement in the application of coding schemes or checking member interviews with the original participants.

Despite my high level of confidence in the validity of the findings and of the conclusions arising from this thesis, readers should bear in mind the above-described caveats.

Implications for Practice

Despite the aforementioned limitations of the research, the findings of this study suggest some clear alternatives in terms of policy directions and practices. First of all, there is a dearth of knowledge about what constitutes a sensible evaluation policy. My research provides a comprehensive review of evaluation policies as well as a framework to facilitate understanding regarding the multidimensionality of the role of evaluation policy for ECB. The 10 categories of evaluation policy identified in this study combined with the understanding of what variables moderate the role of evaluation policy within the organization (as illustrated through the ecological framework) would benefit organizations seeking to develop evaluation policies that are ECB-oriented or are seeking to review their current policies to make them more ECB-friendly. The identification of the variables that moderate the role of evaluation policy in leveraging evaluation capacity can also be used to clarify the types of activities that can be undertaken in order to use evaluation policy as a mechanism for increasing an organization's capacity to do and to use evaluation.

Furthermore, as discussed in the findings, in most of the evaluation policies there is ambiguity around the meaning of learning as a goal for evaluation, as the meaning of learning is generally conflated with the meaning of accountability. Therefore, evaluation policies would benefit from clarifying the meaning of learning and from explicitly linking it

to specific activities and processes that would allow for different types of evaluation utilization. To the extent that the working hypotheses of Cousins and associates, specifically that “data use leads to data valuing” (Cousins et al., 2006; Cousins & Bourgeois, 2014) is supportable, new policies should encourage multiple forms of learning from evaluation so that they are actually used, even if just on a conceptual level or, in the case of process use, for skill building. Cousins & Bourgeois likened the basis to this strategy to Guskey’s (2000) observation regarding educational innovation that “belief follows practice.” Evaluation policies are well positioned to enable leaders to apply a mix of pressure and support to ensure that organization members achieve success in using evaluation (most likely in the form of learning). This will help to integrate evaluation into the organizational culture.

For organization leaders to take an active role in promoting evaluation culture, policymakers need to support these individuals’ evaluation capacity. Training programs that focus on building organization leaders’ capacity and that make direct links to their role in fostering organizational capacity to do and use evaluation are important for the implementation of evaluation policies. In addition to training leaders, policymakers could support the training of other organization members (i.e., other managers and working-level employees) on the virtues of evaluation as an organizational learning system, rather than as an organizational control mechanism. Such initiatives also provide opportunities for some organization members who may be well positioned to become evaluation leaders and champions given their background skills and knowledge (e.g., working with data). Some of the likely consequences of such efforts include increased awareness of the value of evaluation as a learning mechanism within the organization and enhanced organizational capacity for evidenced-based decision making, among other benefits.

Evaluation policies should also encourage the development of indicators and the ongoing monitoring of key organizational performance indicators. Such data should be periodically reviewed and directed to feed into decisions about whether or not to pursue further inquiry (i.e., evaluation), for example. Many of the policies that I reviewed included encouragement of ongoing monitoring but what was unclear, was the extent to which they may have been used. It is imperative that organization members recognize the value and use of indicators; monitoring indicators that are never used are likely to wither on the vine. The implementation of these measures will increase the likelihood of organizational development.

Implications for Research

The findings of this research have expanded and refined the construct of evaluation policy in relation to organizational capacity to do and use evaluation and have added new dimensions to prior theoretical frameworks. This research is indeed breaking new ground in the domain of ECB. We have developed a substantially robust knowledge base and a common set of understandings about what constitutes effective ECB; now it is time to take the next step with this knowledge by situating it within the context of evaluation policy. In order to better understand the role of evaluation policy in ECB, we would benefit from access to a wider range and volume of studies. More research on evaluation policy would provide a more solid basis for understanding variables and patterns that have the potential to provide insight into the mechanisms for capacity development within organizations.

Theoretical Investigations

Our understanding of evaluation policy could be enhanced by more theoretical investigations. Thus far, much of the research work on evaluation policies has focused on the various aspects of evaluation practice that these policies might be used to govern. Findings from this research suggest that the field may benefit from more probing into the ways in which specific areas of evaluation practice interact with the various categories of evaluation policy and how such interaction influences ECB. For instance, how does evaluation policy affect evaluation design? How does it affect ECB activities? How does it affect stakeholders' level of engagement? An interesting way to answer these questions is by undertaking a simulation study that would assign evaluators to two separate contexts—one in which the evaluation would be guided by an evaluation policy and the other in which less guidance would be provided. The evaluators then would be asked to design EBC activities. A comparison may help to identify the influence of the evaluation policy on the planning and design of evaluation activities.

Alternative Methods

I used document analysis, qualitative interviews, and focus groups as the primary means of data collection in this study. Future research might benefit from using quantitative methods to test the variables in the ecological framework in order to provide a different level

of analysis. For example, a survey questionnaire focused on rating the variables moderating the role of evaluation policy could yield some interesting findings in terms of weighting the importance of variables and providing insights into generating new variables.

Another alternative method that could be used comprises case studies of various types of organizations in order to gain a deeper understanding of what evaluation policy means for practical work. Longitudinal case studies could turn out to be quite valuable in capturing the influence of evaluation policy on capacity building over time.

Different Points of View

The ecological framework developed in this research is based on the points of view of the evaluation scholars, community members, and practitioners. Future studies could investigate how the framework resonates with evaluation users, such as, for example, organization or program managers and higher level decision 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.

More Exploration

As was made clear in this research, ECB takes place in complex environments. Thus, more research needs to be undertaken to explore the ways in which particular moderating variables, like leadership, interact with an explicit evaluation policy and how these interactions influence evaluation practice and use.

Finally, as ECB practice continues to expand in the field of evaluation, evaluation experts and community members should engage in more research on evaluation policy in order to further our understanding of the complexities involved in its role in ECB processes and to help answer new questions about current and emerging issues in the field.

Concluding Remarks

Despite the recent calls made by scholars in the evaluation field regarding the importance of evaluation policy and its influence on evaluation practice, there is a lack in the empirical evidence identifying the categories of evaluation policy and explaining its role in building organizational evaluation capacity. In this thesis, I present evidence to fill this

knowledge gap and to support the development of scholarly understanding of the relationship between evaluation policy and organizational capacity for evaluation.

Through the three interconnected studies of my research (a review of an extensive sample of evaluation policies; interviews with scholars and practitioners from Canada, the United States, and Europe; and focus groups with evaluation community members in Jordan and Turkey), the findings of my research make a number of contributions to the evaluation field.

I began the research by developing a conceptual framework that draws on the research of leaders in the evaluation field and acts as a bridge connecting their efforts and expertise. The studies that were most extensively considered were Bourgeois and Cousins' (2008, 2013) empirically grounded framework for organizational evaluation capacity and Trochim's (2009) theoretical framework for evaluation policy. My conceptual framework provides a different and useful lens through which to broaden our understanding of organizational evaluation capacity. When considered together, the two frameworks inform one another, provide a more holistic picture of what to consider when studying organizational evaluation capacity and elucidate the dynamics among the different aspects of the process. I presented the conceptual framework for discussion in a number of conferences in Europe, the United States, and Canada (e.g., the European Evaluation Society Biennial Conference in Dublin, 2014; the American Evaluation Association Conference in Chicago, 2014; and the Canadian Evaluation Association Conference in Montreal, 2015). The framework was well received by those in attendance and generated fruitful discussions about the importance of the subject of the research for the evaluation field.

Over the course of the three successive studies of my research, I (1) validated and refined Trochim's (2009) evaluation policy framework—the only framework on evaluation policy available in the literature—and identified 10 broad categories that capture the goals, processes, strategies, criteria and activities of evaluation policy; (2) revised the one-and-only definition of evaluation policy available in the literature and provided a more comprehensive definition that captures the meaning of the evaluation-policy construct as it is manifested in the extensive sample of evaluation policies included in my research; (3) developed the three-level ecological framework depicting the relationship between evaluation policy and organizational capacity to do and use evaluation and depicting the variables that moderate this relationship at the contextual, organizational and individual levels; and (4) validated the

ecological framework in two countries by means of focus groups conducted in Jordan and Turkey. In my discussion chapter, I developed further understanding of some of the key issues surrounding the role of evaluation policy in fostering organizational evaluation capacity. For example, I discussed the existing tension between learning and accountability—as the two main motives for evaluation—in relation to evaluation policy and provided useful insights aimed at reconciling these motives.

While my research findings reveal that the *current* role of evaluation policy in leveraging organizational evaluation capacity has been limited, my findings also show some evidence that if an evaluation policy is carefully designed to privilege learning as a central and desirable function of evaluation, it will be more likely to have a positive influence on the organizational capacity to do and use evaluation. The passage of time will reveal whether the focus on evaluation policy is a mere transitory trend or instead is a catalyst for transformation in the evaluation field. Given the growing shift in many organizations towards developing and using evaluation policies to set an important frame of reference for the evaluation function, I believe that there does seem to be a tremendous opportunity for evaluators and the evaluation field to use evaluation policies to expand and enhance ECB and the practice of evaluation.

References

- Alkin, M. C., & Ellett, F. (1990). Evaluation models. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation*. Oxford, England: Pergamon Press.
- Anderson, J. E. (1997). *Public policymaking* (3rd ed.). Boston: Houghton Mifflin.
- Anderson, J. E. (2011). *Public policy making* (7th ed.). Boston: Wadsworth, Cengage Learnings.
- Argyris, C., & Schon, D. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison Wesley.
- Arnold, M. E. (2006). Developing evaluation capacity in extension 4-H field faculty: A framework for success. *American Journal of Evaluation*, 27(2), 257–269.
- Armytage, L. (2011). Evaluating aid: An adolescent domain of practice. *Evaluation*, 17(3), 261–276.
- Bardach, E. (2004) Presidential Address -The Extrapolation Problem: How Can We Learn from the Experience of Others? *Journal of Policy Analysis and Management* 23, 205-220.
- Bardach, E., (2006) Policy Dynamics, in: Moran, M., Rein, M., Goodin, R.E. (Eds.), *The Oxford Handbook of Public Policy*. Oxford University Press, New York, USA, pp. 336-366
- Bauer, R. A. (1968). The study of policy formation: An introduction. In R. A. Bauer & K. J. Gergen (Eds.), *The study of policy formation* (pp. 1–26). New York: The Free Press.
- Baxter, P., & Jack, S. (2008). Qualitative study methodology: Study design for novice researchers. *The Qualitative Report*, 13(4), 544–559.
- Bakvis, H. (2000). Rebuilding policy capacity in the era of the fiscal dividend: A report from Canada. *Governance*, 13(1), 71–103.
- Brewer, G. D. & DeLeon, P. (1983). *The foundations of policy analysis*. Homewood, NJ: Dorsey.
- Brinberg, D., & McGrath, J. E. (1985). *Validity and the research process*. Beverly Hills: Sage Publications.
- Brinberg, D., & Hirschman, E.C. (1986). multiple orientations for the conduct of marketing research: an analysis of the academic/practitioner distinction. *Journal of Marketing*, 50, 161–73.

- Bhuyan, A., Jorgensen, A. & Sharma, S. (2010). Taking the Pulse of Policy: The Policy Implementation Assessment Tool. Washington, DC: Futures Group, Health Policy Initiative, Task Order 1.
- 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.
- Bourgeois, I., Toews, E., Whynot, J., & Lamarche, M. K. (2013). Measuring organizational evaluation capacity in the Canadian federal government. *The Canadian Journal of Program Evaluation*, 28(2), 1–19.
- Boyle, R., Lemaire, D., & Rist, R. C. (1999). Introduction: Building evaluation capacity. In R. Boyle & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from practice* (pp. 1–19). New Brunswick, NJ: Transaction.
- Bridgman, P., & Davis, G. (2004). *The Australian policy handbook* (3rd ed.). Sydney, Australia: Allen & Unwin.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1999). Environments in developmental perspective: Theoretical and operational models. In S. L. Friedman & T. D. Wachs (Eds.), *Measuring environment across the life span: Emerging methods and concepts* (pp. 3–28). Washington, DC: American Psychological Association.
- Bronfenbrenner, U. (2005). The developing ecology of human development: Paradigm lost or paradigm regained. In U. Bronfenbrenner (Ed.), *Making human beings human: Bioecological perspectives on human development* (pp. 94–105). Thousand Oaks, CA: Sage.
- Chelmsky, E. (1985) Old patterns and new directions in program evaluation. In E. Chelmsky & American Society for Public Administration. *Program evaluation: Patterns and directions*. Washington, DC: American Society for Public Administration.
- Chelmsky, E. (2009). Integrating evaluation units into the political environment of government: The role of evaluation policy. *New Directions for Evaluation*, 3 (123), 51–66.

- Chelimsky, E. (2012). Valuing, evaluation methods, and the politicization of the evaluation process. *New Directions for Evaluation*, 133, 77–84.
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–36.
- Chouinard, J. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*, 34, 237–253.
- Chouinard, J., & Milly, P. (2016) Mapping the spatial dimensions of participatory practice: A discussion of context in evaluation. *Evaluation and Program Planning*, 54, 1–10.
- Clegg, S. R. (1989). *Frameworks of Power*. London, SAGE
- Colebatch, H. K. (2009). *Policy* (3rd ed.). Buckingham, UK: Open University Press.
- Cousins, J. B., & Bourgeois, I. (2014). Cross-case analysis and implications for research, theory, and practice. In J. B. Cousins & I. Bourgeois (Eds.), *Organizational capacity to do and use evaluation*. *New Directions for Evaluation*, 141, 107–125.
- Cousins, J. B., & Bourgeois, I. (Eds.). (2014). Organizational capacity to do and use evaluation [Special issue]. *New Directions for Evaluation*, 141.
- Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: A review and integration of research-based knowledge*. Charlotte, NC: Information Age.
- Cousins, J. B., Elliott, C., Amo, C., Bourgeois, I., Goh, S., & Lahey, R. (2008). Organizational capacity to do and use evaluation: Results of a pan-Canadian survey of evaluators. *Canadian Journal of Program Evaluation*, 23(3), 1–35.
- Cousins, J. B., & Earl, L. (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London, UK: Falmer.
- Cousins, J. B., & Earl, L. M. (1999). When the boat gets missed: Response to M. F. Smith. *American Journal of Evaluation*, 20(2), 309–317.
- Cousins, J. B., Goh, S., & Clark, S. (2005). Data use leads to data valuing: Evaluative inquiry for school decision making. *Leadership and Policy in Schools*, 4, 155–176.
- Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation*, 19(2), 99–141.
- Cousins, J. B. & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331–364.

- Creswell, J., & Plano-Clark, V. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Cresswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Dablestein, N., & Patton, M. Q. (2013). Lessons learned and the contributions of the Paris Declaration: Evaluation to evaluation theory and practice. *Canadian Journal of Program Evaluation*, 27(3), 173–200.
- Dahler-Larsen, P. (2006) Organizing knowledge: Evidence and the construction of evaluative information systems. In R. C. Rist & N. Stame (Eds.), *From studies to streams: Managing evaluative systems*. Transaction Publishers.
- Daniell, K. A., Morton, A., & Rios Insua, D. (2016). Policy analysis and policy analytics. *Annals of Operations Research*, 236(1), 1-13.
- Datta, L. (2009). Golden is the sand: Memory and hope in evaluation policy and evaluation practice. *New Directions for Evaluation*, 123, 33–50.
- DeLeon, P. (1999). The Stages Approach to the Policy Process: What has it done, Where is it going? In P.A. Sabatier (Ed.), *Theories of the Policy Process*. (pp. 19-34). Boulder, CO: Westview Press
- Ebrahim, A. (2005). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly*, 34(1), 56–87.
- Eisner, E. W. (1979). *The educational imagination: On the design and evaluation*. New York, NY: Macmillan.
- Elmore, R. F (1978). Organizational models of social program implementation. *Public Policy*, 26 (2), 17–31.
- Feinstein, O.N. (2012). Evaluation as a learning tool. *New Directions for Evaluation*, 35, 103–112.
- Forss, K., Rebien, C. C., & Carlsson, J. (2002). Process use of evaluation: Types of use that precede lessons learned and feedback. *Evaluation*, 8(1), 29–45.
- Friedrich, C. J. (1963). *Man and his government*. New York, NY: McGraw-Hill.
- Garcia-Iriarte, E., Suarez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2011). A catalyst for-change approach to evaluation capacity building. *American Journal of Evaluation*, 32(2), 168–182.

- Gaventa, J. (2003). *Power after Lukes: a review of the literature*. Brighton: Institute of Development Studies.
- Gaventa, J. & Cornwall, A. (2006). *Power and knowledge*. 172–89 in Action Research Reader, edited by P. Reason and H. Bradbury. London: Sage.
- Geva-May, I., & Pal, L. A. (1999). Good fences make good neighbours: Policy evaluation and policy analysis—exploring the differences. *Evaluation*, 5(3), 259-277.
- Ghere, G., King, J., Stevahn, L., & Minnema, J. (2006). A professional development unit for reflecting on program evaluator competencies. *American Journal of Evaluation*, 27(1), 108–123.
- Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity Press.
- Goehler, G. (2000). *Types of power relations*. London, SAGE.
- Greenwood, R., & Suddaby, R. (2005). Institutional entrepreneurship in mature fields: the big five accounting firms. *Academy of Management Journal*, 49(1), 27-48.
- Guijt, I. (2010). Accountability and learning: Exploding the myth of incompatibility between accountability and learning. In J. Ubels, N. A. Acquaye-Baddo, & A. Fowler (Eds.), *Capacity development in practice* (pp. 277–291). London, UK: Earthscan.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, 8(3/4), 381–391.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- Howlett, M., & Lindquist, E. (2004). Policy analysis and governance: Analytical and policy styles in Canada. *Journal of Comparative Policy Analysis: Research and Practice*, 6(3), 225–249.
- Howlett, M., & Ramesh, M. (2015). Achilles' heels of governance: Critical capacity deficits and their role in governance failures: The Achilles heel of governance. Regulation & Governance June: <http://dx.doi.org/10.1111/rego.12091>.
- Hornsby, J. S., Kuratko, D. F., Holt, D.T., & Wales, W. J. (2013). Assessing a measurement of organizational preparedness for corporate entrepreneurship. *Journal of Product Innovation Management*, 30(5), 937–995
- Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation*, 29(3), 358–368.

- Jann, W. & Wegrich, K. (2007). Theories of the policy cycle. In F. Fischer, G. J. Miller, M. S. Sidney (Eds.), *Handbook of Public Policy Analysis*. (pp. 43-61). Boca Raton, FL: CRC Press, Taylor & Francis Group
- Jackson, T. (2016). Paternalistic leadership: The missing link in cross-cultural leadership studies. *International Journal of Cross Cultural Management*. 16 (1) 3-7.
- Jacobson, C., Carter, R. W., Hockings, M., & Kelman, J. (2011) Maximizing conservation evaluation utilization. *Evaluation*, 17(1), 53–71.
- Jones, C. O. (1997). *An Introduction to the study of public policy* (3rd ed.). Boston: Wadsworth Publishing Co. Inc.
- Kaufman, H. (1973). *Administrative Feedback*. Washington D.C: The Brookings Institute.
- Knoepfel, P., Larrue, C., Varone, F., & Hill, M. (2011). *Public policy analysis*. Bristol, UK: The Policy Press.
- Kogen, L. (2018). What have we learned here? Questioning accountability in aid policy and practice. *Evaluation*, 24(1), 98–112.
- Labin, S. N., 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(3), 307–338.
- Langton, N., Robbins, S. P., & Judge, T. (2005). *Organizational behavior* (5th ed.). Toronto, Canada: Pearson.
- LaVelle, J., & Donaldson, S. I. (2010). University-based evaluation training programs in the United States 1980–2008: An empirical examination. *American Journal of Evaluation*, 31(1), 9–23.
- Leeuw, F. L. (2009). Evaluation policy in the Netherlands. *New Directions for Evaluation*, 123, 87–102.
- Lehtonen, M. (2014). Evaluating megaprojects: From the “iron triangle” to network mapping. *Evaluation* 20(3), 278–295.
- 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), 12-36
- Leshaem, S., & Trafford, V. (2007). Overlooking the conceptual framework. *Innovations in Education and Teaching International*, 44(1), 93–105.

- Levin-Rozalis, M., Rosenstein, B., & Cousins, B. (2009). A precarious balance: Educational evaluation capacity building in a globalized society. In K. E. Ryan & J. B. Cousins (Eds.), *The Sage international handbook of educational evaluation* (pp. 191–212). Thousand Oaks, CA: Sage.
- Levinson, B. A. U., & Sutton, M. (2001). Introduction: Policy as/in practice—a sociocultural approach to the study of educational policy. In M. Sutton & B. A. U. Levinson (Eds.), *Policy as practice: Toward a comparative sociocultural analysis of educational policy*. Westport, CT: Ablex.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Mansur, J., Sobral, F., & Goldszmidt, R. (2017). Shades of paternalistic leadership across cultures. *Journal of World Business*, 52(5), 702–713.
- Mark, M. M., Cooksy, L. J., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. *New Directions for Evaluation*, 123, 3–11.
- Mathison, S. (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Mayer, D., & Thompson, M. (1982). A case study of policy: Processes, development, and implementation. *Journal of Leisurability*, 9(3), 13–19.
- Mayne, J. (2007). Challenges and lessons in implementing results-based management. *Evaluation*, 13(1), 89–107.
- Mayne, J. (2009). Results management: Can results evidence gain a foothold in the public sector? In O. Reiper, F. Leeuw, & T. Ling (Eds.), *The evidence book* (pp. 117–149). New Brunswick, NJ: Transaction.
- McCool, D. (1995). *Public policy theories, models, and concepts*. Englewood Cliffs, NJ: Prentice Hall.
- McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching people to fish? Building the evaluation capacity of public sector organizations. *Evaluation*, 9, 9–29.
- McGrath, J. E., (1995), Methodology matters: doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction: Toward the Year 2000*, edited by R. Baeker, J. Grudin, W. Buxton, S. Greenberg. San Francisco: Morgan-Kaufmann Publishers, pp. 152–169.
- McNay, L. (1992). *Foucault and Feminism*. Boston: North eastern University Press.
- Meier, K.J., (1999). Are we sure Lasswell did it this way? Lester, Goggin and implementation research. *Policy Currents* 9 (1), 5–8

- Miles, M. S., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Milstein, B., & Cotton, D. (2000). Defining concepts for the presidential strand on building evaluation capacity. Fairhaven, MA: American Evaluation Association.
- 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.
- Nagel, I. H., Freeman, H. E., & Russell Sage Foundation. (1975). Academic and entrepreneurial research: The consequences of diversity in federal evaluation studies. New York, NY: Russell Sage Foundation.
- Neuman, A., Shahor, N., Shina, I., Sarid, A., & Saar, Z. (2013). Evaluation utilization research: Developing a theory and putting it to use. *Evaluation and Program Planning*, 36(1), 64–70.
- 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.
- Norman, R. (2002). Managing through measurement or meaning? Lessons from experience with New Zealand's public sector performance management systems. *International Review of Administrative Sciences*, 68, 619–628.
- Nutley, S., Walter, I., & Davies, H. (2007). *Using evidence: How research can inform public services*. Bristol, UK: Policy Press.
- Office of Management and Budget. (1993). Government Performance and Results Act. Retrieved July 11, 2014, from http://www.whitehouse.gov/omb/mgmt_gpra/gplaw2m
- Office of Management and Budget. (2010). Government Performance and Results Modernization Act. Retrieved July 11, 2014, from <http://www.whitehouse.gov/omb/performance/gprm-act>
- Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report*, 13(4), 695–705.
- O'Toole, Jr., L. J. (2003). Inter organizational relations in implementation. In B. Guy Peters and J. Pierre (eds), *Handbook of Public Administration*, Sage, London, pp. 234–244.
- Payne, C. (2008). *So much reform, so little change: the persistence of failure in urban schools*. Cambridge: Harvard Education Press.

- Parsons, W. (2004). Not just steering but weaving: Relevant knowledge and the craft of building policy capacity and coherence. *Australian Journal of Public Administration*, 63(1), 43–57.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15(3), 311–320.
- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4(2), 225–233.
- Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22(3), 329–336.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2005). Toward distinguishing empowerment evaluation and placing it in a larger context: Take two. *American Journal of Evaluation*, 26, 408–414.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford.
- Peters, G. B. (2005). The Problem of Policy Problems. *Journal of Comparative Policy Analysis: Research and Practice*, 7, 349–370.
- Portz, J. (1996). Problem definitions and policy agendas. *Policy Studies Journal*, 24, 371–386.
- Preskill, H., Zuckerman, B., & Matthews, B. (2003). An exploratory study of process use: Findings and implications for future research. *American Journal of Evaluation*, 24(4), 423–442.
- Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, 35(1), 116–119.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459.
- Pressman, J. L. & Wildavsky, A. B. (1984). *Implementation: how great expectations in Washington are dashed in Oakland* (3rd ed.). Berkely: University of California.
- Princen, S. (2011). Agenda-setting strategies in EU policy processes. *Journal of European Public Policy*, 18, 927–943.
- Rocheftort, D. A., & Cobb, R. W. (1994). *The politics of problem definition: Shaping the policy agenda*. Lawrence: University Press of Kansas

- Rubin, A., & Babbie, E. (2001). *Essential research methods for social work*. Belmont, CA: Brooks/Cole.
- Radin, B. A. (1998). The Government Performance and Results Act (GPRA): Hydra headed monster or flexible management tool? *Public Administration Review*, 58(4), 307–316.
- Raimondo, E. (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, 24(1), 26–41.
- Regeer, B. J., Wildt-Liesveld, R. D., Mierlo, B. V., & Bunders, J. F. (2016). Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation*, 22(1), 6–28.
- Rist, R. C. (2006). The “E” in monitoring and evaluation: Using evaluative knowledge to support a results based management system. In R. C. Rist & N. Stame (Eds.), *From studies to streams: Managing evaluative systems* (pp. 3–22). New Brunswick, NJ: Transaction.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.
- Sabatier, P. A. (Ed.). (1999). *Theories of the policy process*. Boulder, CO: Westview Press.
- Sabatier P. A. & Mazmanian D. (1995). A conceptual framework of the implementation process. In Theodolou, S. Z. & Cahn, M. A. (Eds.), *Public policy: The essential readings*, (pp. 153-173). Upper Saddle River, NJ: Prentice Hall.
- Sanders, J. R. (2002). Mainstreaming evaluation. *New Directions for Evaluation*, 99, 3–6.
- Scriven, M. (1966). *The methodology of evaluation*. Lafayette, IN: Purdue University.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage.
- Segerholm, C. (2003). Researching evaluation in national (state) politics and administration: A critical approach. *American Journal of Evaluation*, 24(3), 353–372.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19, 1–19.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research and practice since 1986. *Evaluation Practice*, 18, 195–208.

- Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice*, 14, 237–242.
- Smith, K. B. & Larimer, C. W. (2009). *The public policy theory primer*. Boulder: Westview Press.
- Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stake, R., & Abma, T. A. (2005). Responsive evaluation. In S. Mathison (Ed.), *Encyclopaedia of evaluation* (pp. 376–379). Thousand Oaks, CA: Sage.
- Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: The Guilford Press.
- Stern, E. (2009). Evaluation policy in the European Union and its institutions. *New Directions for Evaluation*, 123, 67–85.
- Stevahn, L., King, J. A., Ghore, G., & Minnema, J. (2005). Evaluator competencies in university-based training programs. *Canadian Journal of Program Evaluation*, 20(2), 101–123.
- Stockdill, S. H., Baizerman, M., & Compton, D. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7–25.
- Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (2000). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA: Kluwer Academic.
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation* (120), 55–69.
- Telfair, J., & Leviton, L. C. (1999). The community as client: Improving the prospects for useful evaluation findings. *New Directions for Evaluation*, 83, 5–16.
- Thomas, V. G. (2004). Building a contextually-responsive evaluation framework: Lessons from working with urban school interventions. *New Directions for Evaluation*, 101, 3–23.
- Titmuss, R. M. (1974). *Social policy*. London, UK: Allen & Unwin.
- Treasury Board of Canada. (2009). *Policy on evaluation*. Retrieved July 10, 2014, from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=15024>
- Treasury Board of Canada. (2010). *About the Centre of Excellence for Evaluation*. Retrieved July 10, 2014, from <http://www.tbs-sct.gc.ca/cee/index-eng.asp>

- Trochim, W. (2008). *Evaluation policy and evaluation practice for large scientific research initiatives*. Presentation to the Annual Conference of the American Evaluation Association, Denver, Colorado, November 6, 2008.
- Trochim, W. (2009). *Evaluation policy and evaluation practice: Taxonomy and methodology*. Presentation to the Annual Conference of the American Evaluation Association, Orlando, Florida, November 14, 2009.
- Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. *New Directions for Evaluation*, 123, 13–32.
- VeneKlasen, L., Miller, V., (2002) *Power and empowerment*. PLA Notes, 43: 39-41
- VeneKlasen, L. & Miller, V. (2002). *A New Weave of Power, People and Politics: The Action Guide for Advocacy and Citizen Participation*. Oklahoma City: World Neighbors IDS Bulletin.
- Vo, A. T. (2013). Visualizing context through theory deconstruction: A content analysis of three bodies of evaluation theory literature. *Evaluation and Program Planning*, 38, 44–52.
- Vo, A. T., & Christie, C. A. (2015). Advancing research on evaluation through the study of context. *New Directions for Evaluation*, 148, 43–55.
- Shadish, W. R. Jr., Cook, T. D., & Leviton, L. C. L. (1991). *Foundations of program evaluation*. Thousand Oaks, CA: Sage.
- Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice*, 14(3), 237–242.
- Weaver, L., & Cousins, J. B. (2004). Unpacking the participatory process. *Journal of Multidisciplinary Evaluation*, 1, 19–40.
- Weiss, C. H. (1973). The politics of impact measurement. *Policy Studies Journal*, 1(3), 179–183.
- Weiss, C. H. (1977). *Using social research in public policy making*. Lexington, MA: Lexington Books.
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426–431.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21–33.

Williams, D. (2001). Issues and practices related to mainstreaming evaluation: Where do we flow from here? *New Directions for Evaluation*, 99, 63–83.

Wilkin, P., & Boyle, ?? (2011). Standards and qualities. In J. Lonsdale, P. Wilkins, & T. Ling (Eds.), *Performance audit: Contributing to accountability in democratic government*, 147–173. Cheltenham, UK: Edward Elgar.

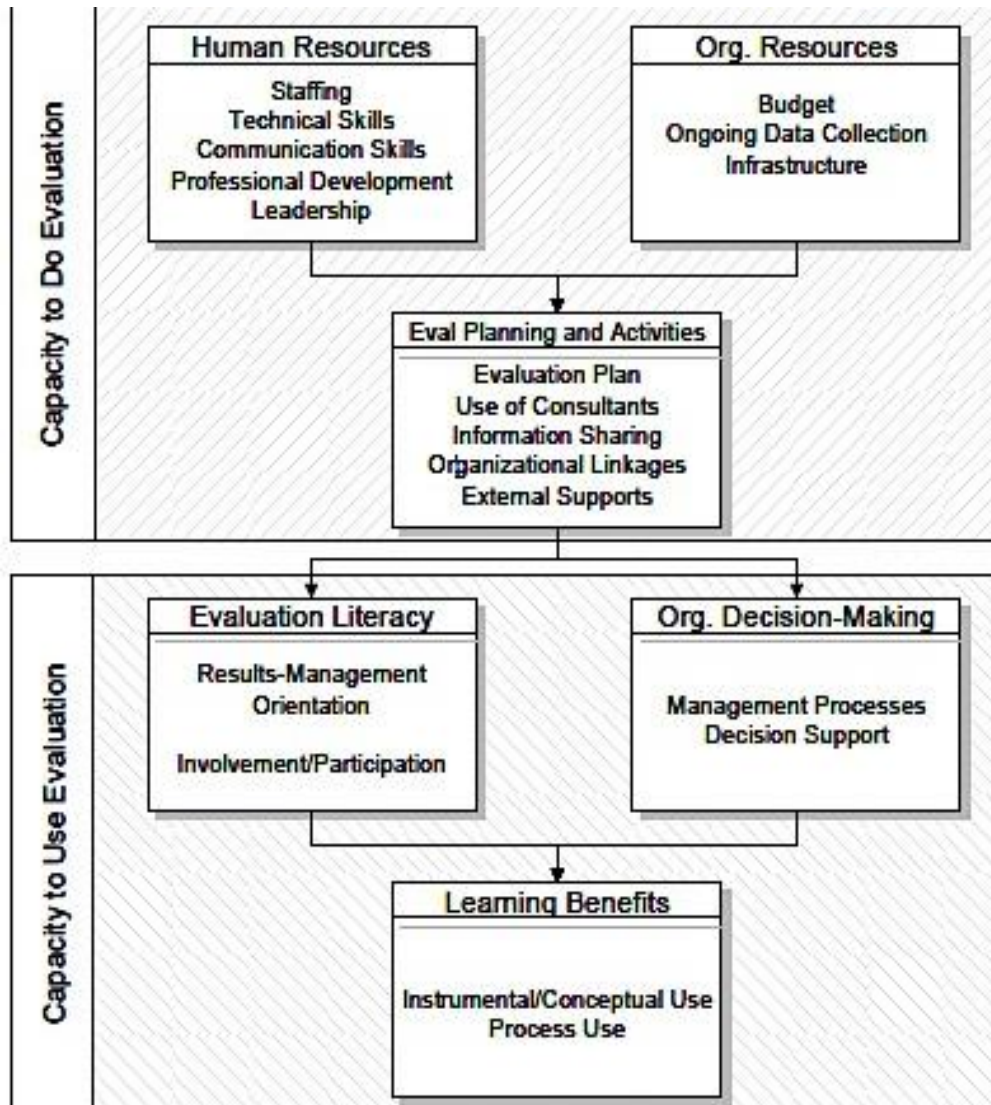
Worthen, B. R. (1990). Program evaluation. In H. J. Walberg & G. D. Haertel (Eds.), *International encyclopaedia of educational evaluation*. New York, NY: Pergamon.

Worthen, B. R. (2001). Whither evaluation? That all depends. *American Journal of Evaluation*, 22, 409–418.

Appendices

Appendix A: Dimensions and Sub-dimensions of Organizational-Evaluation Capacity

Dimensions and sub-dimensions of organizational-evaluation capacity (Bourgeois & Cousins, 2008, p. 136)



Appendix B: List Evaluation Policies

Multi-lateral Organizations

African Development Bank (AFDB):

<http://beta.afdb.org/sites/default/files/documents/idev/African%20Development%20Bank%20Independent%20Evaluation%20Strategy%202013%20-%202017.pdf>

<http://www.afdb.org/fileadmin/uploads/afdb/Documents/Evaluation-Reports/18854239-EN-INDEP-EVALU-POLICY-AND-FUNCTIO-RESPO-OF-THE-OPEV.PDF>

Asian Development Bank

<http://www.adb.org/sites/default/files/institutional-document/32104/r297-08.pdf>

Economic Commission for Latin America and the Caribbean (ECLAC):

<http://www.unevaluation.org/document/detail/1200>

Economic and Social Commission for Western Asia (ESCWA):

<http://css.escwa.org.lb/pptcd/EvaluationPolicy.pdf>

International Labour Organization (ILO):

http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_168289.pdf

International Federation of Red Cross and Red Crescent Societies (IFRC):

<https://www.ifrc.org/Global/Publications/monitoring/IFRC-Framework-for-Evaluation.pdf>

<http://www.ifrc.org/Global/Publications/monitoring/IFRC-ME-Guide-8-2011.pdf>

International Fund for Agricultural Development (IFAD):

<http://www.ifad.org/gbdocs/eb/102/e/EB-2011-102-R-7-Rev-2.pdf>

<http://www.ifad.org/pub/policy/oe.pdf>

Office for the Coordination of Humanitarian Affairs (OCHA):

<http://www.unevaluation.org/document/detail/1199>

Organization for Economic Cooperation and Development (OECD):

<http://www.oecd.org/derec/guidelines.htm>

<http://www.oecd.org/development/evaluation/qualitystandards.pdf>

The MDG Achievement Fund (MDG-F):

<http://www.unevaluation.org/document/detail/1198>

United Nations Children's Fund (UNICEF):

[http://www.unicef.org/about/employ/files/2013-14--
Final approved Revised Evaluation Policy.pdf](http://www.unicef.org/about/employ/files/2013-14--Final_approved_Revised_Evaluation_Policy.pdf)

United Nations Development Program (UNDP):

<http://web.undp.org/evaluation/documents/evaluation-policy.pdf>

United Nations Educational Scientific and Cultural Organization (UNESCO):

<http://unesdoc.unesco.org/images/0023/002322/232246e.pdf>

United Nations Environment Program (UNEP):

<http://www.unevaluation.org/document/detail/1157>

United Nations Population Fund (UNFPA):

[http://www.unfpa.org/admin-resource/executive-board-united-nations-development-
programme-united-nations-population-fund-1](http://www.unfpa.org/admin-resource/executive-board-united-nations-development-programme-united-nations-population-fund-1)

United Nations Industrial Development Organization (UNIDO):

[http://www.unido.org/fileadmin/user_media_upgrade/Resources/Evaluation/UNIDO_Ev
aluation_Policy_UNIDO-DGB-M-98-Rev-1_150319.pdf](http://www.unido.org/fileadmin/user_media_upgrade/Resources/Evaluation/UNIDO_Evaluation_Policy_UNIDO-DGB-M-98-Rev-1_150319.pdf)

United Nations Institute for Training and Research (UNITAR):

[http://www.unitar.org/sites/default/files/infocus/monitoring-evaluation-policy-
framework.pdf](http://www.unitar.org/sites/default/files/infocus/monitoring-evaluation-policy-framework.pdf)

United Nations High Commissioner for Refugees (UNHCR):

<http://www.unhcr.org/3d99a0f74.html>

United Nations Development Fund for Woman (UNIFEM):

<http://www.unevaluation.org/document/detail/440>

World Conservation Union (IUCN):

http://cmsdata.iucn.org/downloads/handbook_eng.pdf

https://cmsdata.iucn.org/downloads/evaluation_policy_eng.pdf

World Health Organization (WHO):

http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf

World Intellectual Property Organization (WIPO):

[http://www.wipo.int/export/sites/www/about-
wipo/en/oversight/iaod/evaluation/pdf/2010_2015_evaluation_strategy.pdf](http://www.wipo.int/export/sites/www/about-wipo/en/oversight/iaod/evaluation/pdf/2010_2015_evaluation_strategy.pdf)

[http://www.wipo.int/export/sites/www/about-
wipo/en/oversight/iaod/evaluation/pdf/evaluation_guidelines.pdf](http://www.wipo.int/export/sites/www/about-wipo/en/oversight/iaod/evaluation/pdf/evaluation_guidelines.pdf)

Bilateral Organizations:

Austrian Development Agency:

http://www.entwicklung.at/uploads/media/Guidelines_ProjectProgramme_Evaluations_Juli2009_01.pdf

Australian Aid Agency:

<http://www.oecd.org/dac/evaluation/Ausaidevalguidelines.pdf>

Canadian International Development Agency (CIDA):

<http://www.oecd.org/dac/evaluation/canada.pdf>

[http://www.acdi-cida.gc.ca/inet/images.nsf/vluimages/evaluation3/\\$file/rolling-five-year-2013-14-to-2017-18-eng.pdf](http://www.acdi-cida.gc.ca/inet/images.nsf/vluimages/evaluation3/$file/rolling-five-year-2013-14-to-2017-18-eng.pdf)

Dutch Ministry of Foreign Affairs:

<http://www.iob-evaluatie.nl/sites/iob-evaluatie.nl/files/000%20Evaluation%20Policy%20and%20Guidelines%202009.pdf>

French Development Agency:

http://www.afd.fr/webdav/site/afd/shared/RECHERCHE/Evaluation/document/AFD_evaluation_policy.pdf

The Fogarty International Center (FIC):

<http://www.fic.nih.gov/about/staff/policy-planning-evaluation/pages/evaluation-framework.aspx>

New Zealand Aid Program:

<http://www.aid.govt.nz/sites/default/files/Evaluation%20Policy.pdf>

Swedish International Development Cooperation Agency (SIDA):

http://betterevaluation.org/sites/default/files/SIDA3753en_Looking_back_0.pdf

Swiss Agency for Development and Cooperation (SDC):

https://www.eda.admin.ch/content/dam/deza/en/documents/Effectiveness/resource_en_181530.pdf

UK Department for International Development (DFID):

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/204119/DFID-Evaluation-Policy-2013.pdf

<https://www.gov.uk/government/publications/dfid-evaluation-policy-2013>

USAid:

<https://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf>

International Development Research Centre (IDRC):

<http://www.idrc.ca/EN/Documents/Evaluation-at-IDRC.pdf>

The Spanish Agency for International Development Cooperation:

[http://www.exteriores.gob.es/Portal/es/SalaDePrensa/Multimedia/Publicaciones/Documents/INT%20SPANISH%20COOPERATION%20\(on-line\)2indd.pdf](http://www.exteriores.gob.es/Portal/es/SalaDePrensa/Multimedia/Publicaciones/Documents/INT%20SPANISH%20COOPERATION%20(on-line)2indd.pdf)

Appendix C: Evaluation Policy Data Elaboration

Document	Context	Purpose of the Policy	Main Focus
The United States Agency for International Development (USAID)			
USAID Evaluation Policy (2011)	USA International development	To set out requirements, standards and practices to address contemporary needs in response to increasingly complex operating environments and to address the demands of strong practice and use of evaluation as a crucial tool to inform evidence-based decision making	The policy was developed in response to calls from within USAID with the intention of learning more systematically from their work and of rigorously and credibly documenting their programs' effectiveness. The policy is a framework for organizational learning with consideration for continuing to make improvements and adjustments in order to incorporate new ideas and maintain policy relevance.
USAID Evaluation Policy (2016)		To provide clarity to staff, partners and stakeholders regarding the purposes of evaluation, the types of evaluations that are required, and the approaches for conducting, disseminating and using evaluations. It also serves to communicate to implementing partners and key stakeholders a new approach to evaluation.	This policy helps to implement the <i>Foreign Aid Transparency and Accountability Act</i> of 2016 for USAID and works in concert with existing Agency policies, strategies and operational guidance, including those related to project design, staff members' evaluation-related competencies, performance monitoring, knowledge management and research management.

Document	Context	Purpose of the Policy	Main Focus
The World Intellectual Property Organization (WIPO)			
WIPO Evaluation Policy (2010)	WIPO works with 184 Member States to promote intellectual property around the globe.	To describe the concept and role of evaluation within WIPO; the roles and responsibilities of the program and of the project managers; and the prioritization, planning, management, conduct, budgeting, follow-up and use of evaluations.	This Policy provides a comprehensive framework for planning and conducting independent evaluations as well as for informing policy and strategic decisions with evaluation evidence with a view to improving the development impact of WIPO's activities. The Policy aims to enhance the generation and use of value-added evaluative information for (i) decision-making processes related to the improvement of present and future activities; (ii) policy formulation and review by Member States; and (iii) management oversight by the Director General. The Policy emphasizes the fact that evaluation is an organizational responsibility focused on learning and accountability and the fact that it applies to independent evaluations conducted throughout the organization.
WIPO Evaluation Policy Second Edition / 2016-2020		To provide the framework for planning and conducting evaluations within WIPO and for reporting on and using the results from such evaluations. The Policy establishes principles for the evaluation function within WIPO's program-cycle management system.	The Policy aims to: (a) set the conditions for the evaluation function to support learning and accountability and to help promote effective practices and results-based program-cycle management within WIPO, of which evaluation is an integral part; (b) define principles for the conduct of evaluation activities and their use for WIPO and its stakeholders; and (c) define the positioning and main tasks of the evaluation function.

Document	Context	Purpose of the Policy	Main Focus
Treasury Board of Canada Secretariat			
Policy on Evaluation (2009)	Canadian government	To clarify the role of evaluation within the government’s management framework to ensure that the government has timely, strategically focused, objective and evidence-based information on the performance of its policies, programs and initiatives	The policy provides clear expectations and standards for evaluation within the Treasury Board Secretariat as well as for evaluations that involve partnerships with other federal organizations, with other levels of government, and/or with the private or not-for-profit sectors.
Policy on Results (2016)		To improve the achievement of results across government and to enhance the level of understanding of the results that the government seeks to achieve and does achieve, as well as of the resources used to achieve those results	This policy replaces three Treasury Board policies: Policy on Reporting of Federal Institutions and Corporate Interests to Treasury Board of Canada Secretariat (2007); Policy on Evaluation (2009); and Policy on Management, Resources and Results Structures (2010). It clarifies requirements, responsibilities and expectations related to the use of the results of performance measurement to manage and improve programs, policies and services and related to the resources used to do so.
United Nations Development Program (UNDP)			
The evaluation policy of UNDP (2006)	International development	To establish a common institutional basis for the UNDP evaluation function; to increase transparency, coherence and efficiency in generating and using evaluative	This policy establishes the guiding principles and norms; explains key evaluation concepts; outlines the main organizational roles and responsibilities; defines the types of evaluation covered; and identifies the key elements of a

Document	Context	Purpose of the Policy	Main Focus
		knowledge for organizational learning and effective management for results; and to support accountability	system for learning and knowledge management. It also outlines the capacity and resource requirements to enhance excellence in the development of an evaluation culture and a learning oriented organization.
The evaluation policy of UNDP (2016)		To set out the purpose and basic principles of evaluation and define the institutional architecture for UNDP and its associated funds and programs; to cover the independent evaluations conducted by the Independent Evaluation Office of UNDP, the decentralized evaluations, and the activities of UNDP and the Independent Evaluation Office in support of national evaluation capacity	The evaluation policy is revised to align with the overall mandates of UNDP and its associated programs and funds, and with the Charter of the United Nations and its objectives. It focuses on capacity building, quality assurance and policy implementation.
United Nations Evaluation Group (UNEG)			
Standards for Evaluation (2005)	UNEG is an inter-agency professional network that brings together the evaluation units of the United Nations system, including United Nations departments, specialized agencies, funds, programs and affiliated organizations.	To define norms and standards that aim at contributing to the professionalization of the evaluation function as a step to making further progress in a system-wide collaboration on evaluation, in particular in the harmonization and simplification of methodologies, norms, standards and cycles of evaluation	The standards are drawn from the best practices of UNEG members and are intended to guide the establishment of the institutional framework, the management of the evaluation function, and the conduct and use of evaluations. They are also a reference for the competencies of evaluation practitioners and work ethics.

Document	Context	Purpose of the Policy	Main Focus
Norms and Standards for Evaluation (2016)		To respond to increasing demands for accountability in evaluation; to ensure continued relevance in guiding and further strengthening evaluation practices; and to serve as the framework for the UNEG's peer reviews and evaluation competency and benchmarking initiatives	The document provide a definition of evaluation, state evaluation's purpose and role within the organization, and outline general and institutional norms for evaluation as well as standards for the institutional framework for evaluation, for the conduct and management of evaluation, for evaluation competencies, and for evaluation quality. All UNEG members are required to implement the norms and standards outlined in the document in order to bring a consistent and harmonized approach to the continual improvement of the UN evaluation system over time.
United Nations Industrial Development Organization (UNIDO)			
UNIDO evaluation policy (2006)	International organization mandated to promote industrial development for poverty reduction, inclusive globalization and environmental sustainability	To recognize the importance of an efficient and comprehensive evaluation system aligned with international policies, standards and practices, for measuring the impact of UNIDO's work at field level	The Policy focuses on the importance of evaluation for management, and serves as an input to provide decision-makers with knowledge and evidence about performance and practices. Evaluation users are project managers, team leaders and UNIDO management at all levels, working in both the technical branches and country offices, at headquarters and in the field, as well as UNIDO counterparts and donors. Evaluation recommendations must be practical and evaluation users should participate actively in the entire evaluation process.
UNIDO evaluation policy (2015)		To reflect in the updated Policy the modifications resulting from the issuance of another policy and recent changes within the	The new roles of the Director General and the Executive Board are to ensure that evaluation contributes to accountability, supports management decision-making, and facilitates

Document	Context	Purpose of the Policy	Main Focus
		organization (i.e., the renaming of the Office and the discontinuation of its participation in the appraisal and approval committees for programs and projects)	the implementation of recommendations and the dissemination of lessons learned from evaluations throughout the entire organization.
United Nations Educational, Scientific and Cultural Organization (UNESCO)			
UNESCO Evaluation Policy (2008)	UNESCO is responsible for coordinating international cooperation in education, science, culture and communication and for strengthening the ties between nations and societies in these areas.	To provide definitions, concepts, criteria, principles, roles and responsibilities, and a priority-setting model in order to identify the external evaluations that should be undertaken	The policy responds to the Director-General's request that UNESCO elaborate on its evaluation policy and strategy taking into account the views expressed by Member States—and, in particular, on (i) the need to include criteria for selecting activities, themes and programs for evaluation; and (ii) policy statements on evaluation with a clear definition of the roles and responsibilities of each sector or division concerned outlining the necessary interaction and cooperation between auditing, evaluation, programming and program monitoring. The policy document includes: evaluation definition, evaluation criteria, key principles when undertaking evaluation, roles and responsibilities, the evaluation priority-sitting model, and resource requirements.
UNESCO Evaluation Policy (2014-2021)		The reviewed policy reflects changes that have taken place both internal to UNESCO as well as externally. It signifies rising expectations regarding the transparency, accountability and	This policy establishes a framework for strengthening UNESCO's overall evaluation system based upon the principles of independence, accountability, transparency, utility and impartiality in its evaluation practices; methodological rigour throughout;

Document	Context	Purpose of the Policy	Main Focus
		effectiveness of the work of UNESCO and the efficient use of resources allocated for its key mandates.	and an orientation towards reinforcing organizational and national evaluation capacities. It is based on: (1) emphasis on knowledge generation, program improvement and organizational learning; (2) a strong corporate culture of accountability for achieving results and for using resources efficiently; (3) an integrated evaluation system; and (4) a renewed commitment to support and use evaluation
International Union for Conservation of Nature (IUCN)			
Managing Evaluation: A Guide for IUCN Programs and Managers (2004)	International development involves states, government agencies and non-governmental organizations.	To promote evaluation as a formal activity within the organization and to assist IUCN project and program managers who initiate and manage evaluation activities	The guide provides a framework for establishing a Monitoring & Evaluation System for IUCN that: supports learning, improvement and accountability through regular reviews of the relevance, effectiveness, efficiency and impact of IUCN's work at the project, program and organizational levels; promotes a learning culture of self-assessment, reflection and internal review as well as external review; and provides training and capacity building for IUCN managers in evaluation and self-assessment. The document draws on materials and policies developed by other organizations (e.g., the World Bank).
The IUCN Monitoring and Evaluation Policy (2013)		To demonstrate IUCN's commitment to monitoring and evaluating its work, to using the results to drive performance and impact, and to setting out minimum evaluation requirements, principles,	The policy establishes common structures and standards across the IUCN to contribute to the development of mechanisms for information sharing, coordination and alignment within the organization. It differentiates between monitoring and evaluation and their roles and

Document	Context	Purpose of the Policy	Main Focus
		and roles and responsibilities, while also providing external stakeholders, such as donors, IUCN members, partners and users, with information on IUCN standards and procedures	uses. The results of the M&E activities are used for management and decision-making process.
Economic Commission for Latin America and the Caribbean (ECLAC)			
Preparing and Conducting Evaluations: ECLAC Guidelines (2005)	Regional economic and social development of the Latin American and Caribbean countries (46 member states)	To strengthen the organization's system of monitoring and evaluation in order to more accurately and effectively measure the impact of its work and in order to respond to the increased focus on the importance of monitoring and evaluation for results-based management frameworks in the wider UN system	These guidelines establish an evaluation structure that outlines evaluation requirements, processes, roles and responsibilities in order to ensure accountability, support results-based management systems, and promote learning and innovation. Evaluation is seen as a means of contributing to the credibility and legitimacy of ECLAC's participation in the region's economic and social development processes.
ECLAC Evaluation Policy and Strategy (2014)		To maximize transparency and coherence, ensure high quality standards in evaluations, and contribute to greater accountability, improved performance and improved institutional learning	The policy emphasizes the importance of evaluation as an integral part of the work of ECLAC in all of its substantive areas. It is an accountability mechanism to enhance ECLAC's legitimacy and credibility, to ensure a results orientation in the activities that it undertakes, and to identify the impacts of its interventions. It is a key driver of institutional learning.

Document	Context	Purpose of the Policy	Main Focus
Swedish International Development Cooperation Agency (SIDA)			
Sida's Evaluation Policy (1999)	Sida is a government agency that works on behalf of the Swedish parliament and government, with the mission of reducing poverty in the world.	To promote a comprehensive evaluation function within the organization and to identify objectives, roles and responsibilities, evaluation criteria and procedures for evaluation management. The policy is focused on developing a comprehensive and strategic evaluation system.	Sida works in accordance with the directives of the Swedish parliament and government in conducting its development work. The organization developed the evaluation policy to guide its sectoral and regional departments and to identify their roles and responsibilities. The policy objectives are to increase control (accountability) and promotion (learning).
Sida's Evaluation Policy (2007)		To develop existing capacities and work practices, including the examination of the need to develop Sida's ability and that of Sweden's cooperation partners to deal more efficiently with evaluation capacity development efforts. This policy is the result of a review of Sida's evaluation function and of its current priorities for evaluation and capacity development.	The policy shows Sida's renewed commitment to capacity development and responds to its accountability requirements. The policy reflects changes in the organizational context and in the contexts of the various partner countries as a result of changes in their laws, policies, norms and values. It indicates that the strengthening of the human-rights framework and the enhancement of understanding of and respect for human rights are key concerns in Swedish development cooperation. The new policy includes a step-by-step guide for Sida program officers and for others who are involved in the management of evaluations initiated by Sida or its partners.

Document	Context	Purpose of the Policy	Main Focus
United Nations International Children's Emergency Fund (UNICEF)			
UNICEF evaluation policy (2008)	International development: UNICEF is active in 192 countries, territories, and other areas	To foster a common institutional understanding of the evaluation function within UNICEF and to strengthen evidence-based decision-making and advocacy, transparency, coherence and effectiveness	This policy was required to assess the performance of UNICEF against the organization's strategic plan. It aims to ensure that UNICEF has access to timely, strategically focused and objective information on the performance of its policies, programs and initiatives in order to produce better results for stakeholders at the global, regional and country levels.
Revised Evaluation Policy of UNICEF (2008)		To govern the organization's evaluation function and provide a comprehensive framework for all evaluation activities undertaken by UNICEF, addressing not only the conduct of evaluations and the use of evaluation results, but also the development of the evaluation function within UNICEF and its role in strengthening evaluation capacity among national partners	UNICEF has revised its Evaluation Policy in order to respond to changes within the organization and in its operating environment and to rising expectations regarding the transparency, accountability and effectiveness of development processes and regarding the efficient use of resources allocated for development. Areas for improvement were identified while implementing the previous policy. Evaluation planning needed to be strengthened and linked to the organization's strategic objectives, and provision of human and financial resources to evaluation needed greater attention. Changes in the programming environment have resulted in increased demand for formative evaluation to inform ongoing policy, strategy and program development, as well as in calls for more impact evaluation to gauge results. Further, there is growing interest in strengthening the evaluation capacity of national partners.

Document	Context	Purpose of the Policy	Main Focus
African Development Bank (AFDB)			
Independent Evaluation Policy (2007)	The African Development Bank is engaged in promoting economic development and social progress throughout Africa.	To establish an evaluation function within the organization that is able to resolve issues caused by conflicts of interest in the ideas in which the AFDB is involved, in particular by clarifying disclosure procedures so that all employees/board members know what interests must be divulged, as well as by articulating definitions, concepts, criteria, principles, processes, and roles and responsibilities	The policy was created as a response to the recommendation of the Committee of Operations and Development Effectiveness (CODE) to develop an evaluation policy in order to address the issue of conflicts of interest related to evaluation. It establishes the framework for the evaluation function and includes measures to avoid conflicts of interest that are applicable to both staff and consultants. The policy also presents evaluation operational procedures including organizational measures to integrate the evaluation function and enhance its effectiveness.
24-Revised Independent Evaluation Policy (2016)		To reflect new updates as a result of the assessment of the evaluation function and as a result of changes in priorities while placing a particular emphasis on enhancing the development effectiveness of the organization through independent evaluations, oversight of self-evaluation processes and products, and proactive engagement in evaluation partnerships and knowledge-sharing activities	The revised policy draws on the findings of an internal assessment of the organization's evaluation function undertaken 10 years after the creation of the first evaluation policy and provides an updated governance framework for this function. The policy takes into account the new Bank Group's core priorities as defined in the Ten Year Strategy, the evolution of evaluation practices within other multilateral development banks and in the development evaluation community, and regional member countries' greater ownership of the development process and intensified focus on results. The new policy is in line with internationally accepted standards, including the Development Assistance Committee's Principles for Evaluation of Development

Document	Context	Purpose of the Policy	Main Focus
			Assistance and the Good Practice Standards issued by the Evaluation Cooperation Group.
UK Department for International Development (DFID)			
The UK's policy on evaluation for international development (2009)	DFID is the UK government department responsible for managing international development aid.	To provide a framework for DFID's independent evaluation of the UK's contribution to international development, and to set standards that are to be applied to all evaluations undertaken. The policy aims to strengthen institutional structures and to ensure greater results focus, accountability and quality assurance in DFID's work.	The policy reflects changes in the framework for providing international development assistance as a result of the Paris Declaration on Aid Effectiveness 2005. The UK developed this policy to accelerate the actions required to meet the targets set for 2010 to implement the Paris commitments. The policy puts in place procedures and measures to strengthen the quality of DFID's monitoring and information systems and those of its partners, particularly in developing countries, as well as to increase mutual accountability and foster capacity development.
The UK's policy on evaluation for international development (2013)		To raise awareness of the importance of evaluating development spending to determine whether the desired effects are being achieved. It provides clarity and consistency in the design, conduct and use of evaluation. It also sets high standards, rigour and ethical practices in evaluation to ensure its quality and utility.	The revised policy was needed to meet new transparency and independence requirements that resulted from the 2011 Independent Commission for Aid Impact (ICAI) which requires direct reporting to Parliament on the results and value for money of all official development assistance. The policy is updated to ensure consistency in the application of principles and standards in evaluation.

Document	Context	Purpose of the Policy	Main Focus
Canadian International Development Agency (CIDA)			
CIDA Evaluation Guide (2004)	CIDA was a Canadian federal-government organization that administered foreign aid programs in developing countries. The agency was merged into Global Affairs Canada in 2013 but carried forward its programs and work. The agency had an evaluation plan that was transferred to Global Affairs Canada without major changes.	To build understanding of the evaluation function and of the role of evaluators; to enhance the value realized from CIDA's investments in evaluations; to ensure compliance with Treasury Board Secretariat (TBS) requirements; and to promote consistent management and work practices, both at headquarters and in the field	The document is very detailed. It focuses on how evaluations of CIDA's investments in development cooperation are to be carried out and what they are expected to achieve to make a meaningful and enduring contribution. It sets out the process to be followed, acceptable standards of performance and appropriate work practices. It is designed to promote and facilitate informed decision-making throughout the evaluation process, through the following succession of stages: planning/design, implementation, the reporting and sharing of results. The document includes a "model text" for the evaluation plan, the evaluation report and the executive summary.
Development Evaluation Plan (2015-2020) Global Affairs Canada	Global Affairs Canada is the Canadian government department that manages international development and humanitarian assistance.	To align with and support the departmental management, resources and results structure, and to support the requirements of the Expenditure Management System, including spending reviews	This recent evaluation plan was developed following an in-depth review of the organization's evaluation function in order to meet the requirements of the <i>Financial Administration Act</i> and of the 2009 Treasury Board Policy on Evaluation. It focuses on accountability, knowledge dissemination, and the influence on policy and program decision-making. The plan identifies the evaluation division of responsibilities in order to ensure the development and implementation of a systematic approach that provides technical support (including training and tools) for branch-led evaluations. The plan recognizes the need to optimize the use of evaluations and indicates that a new Standard Operating

Document	Context	Purpose of the Policy	Main Focus
			Procedure for finalizing evaluations through increased engagement with the department and for increasing the use of evaluation findings, recommendations and lessons learned will be implemented. A Report on Lessons Learned from Development Evaluations will be produced in order to promote the use of evaluation knowledge and ensure its wide dissemination.
Danish Development Cooperation			
Evaluation Policy for Danish Development Cooperation (2016)	Danish Development Cooperation is Denmark's government development agency. It provides humanitarian aid and development assistance to developing countries.	To promote accountability and learning; to provide clarity and transparency in the conduct of evaluations of cooperation with developing countries; to further a shared understanding among stakeholders of the priorities, usefulness and value-added of evaluations; to outline principles and standards of quality and utility of evaluations.	The organization originally had "evaluation guidelines" but eventually realized the need to have an actual evaluation policy to strengthen the evaluation function and increase accountability and learning. The policy articulates the purposes, principles and main procedures for an evaluation. It focuses on transparency around evaluation processes and emphasizes evaluation quality and utility. It highlights the importance of capacity building and of learning from evaluation by promoting real-time evaluation as an approach to responding to contextual challenges and by making changes to programs during implementation in order to achieve the planned results.
Evaluation Guidelines for Danish Development Cooperation (2018)		To provide operational guidance for evaluation and to complement the Evaluation Policy (2016).	The guidelines include details about all stages of evaluation including scoping, designing, managing, assigning, implementing, reporting, following up and disseminating evaluation

Document	Context	Purpose of the Policy	Main Focus
			results. The document explains the new evaluation policy; it was created to ensure proper implementation of the policy. The document emphasizes learning and capacity building processes as critical aspects of all evaluations and provides concrete steps to ensure the use of evaluation findings and the implementation of recommendations. These guidelines replace the previous evaluation guidelines published 2012.
World Health Organization (WHO)			
WHO Evaluation Guidelines (2006)	WHO is an international development organization concerned with the protection and advancement of public health, working through offices in more than 150 countries and supporting their national health policies and strategies.	To provide technical references that can clarify what evaluation is, to promote evaluation culture, and to increase staff knowledge of evaluation in order to facilitate collaboration with partners	The document contains definitions of evaluation terms, norms, and standards for evaluation, and guidance on planning, preparing and using evaluation. The main focus of the document is on understanding evaluation, the basic requirements of a credible evaluation, and quality standards. It does not include elements related to capacity building or information about evaluation budgeting and resources.
WHO Evaluation Policy (2012)		To define the overall framework for evaluation, to foster evaluation culture and evaluation use, and to facilitate the conformity of evaluation with best practices and with the norms and standards for evaluation of the United Nations Evaluation Group	The policy indicates that the previous 2006 evaluation guidelines did not succeed in fostering an evaluation culture, in developing evaluation capacity or in promulgating participatory approaches to evaluation. For this reason, the organization established this policy to create a mechanism for evaluation that would ensure accountability for performance and results, and would reinforce learning and the utility and quality of

Document	Context	Purpose of the Policy	Main Focus
			evaluation. The new policy includes information about financing evaluation and about roles and responsibilities in order to ensure proper planning, management, conduct and use of evaluation.
International Labour Organization (ILO)			
A new policy and strategic framework for evaluation at the ILO (2004)	The ILO is an international development organization that works with governments and employers to set labour standards and develop policies.	To improve and strengthen the practice of independent evaluation, and to establish principles for systematic self-evaluation of program performance	The development of this policy was supported by the creation of a central evaluation unit to ensure the implementation of the policy and to submit evaluations directly to the Director-General. The policy focuses on accountability and transparency and sets clear guidelines for planning, undertaking and using evaluation that contribute to sound decision-making, optimize resource allocation and improve management.
ILO policy guidelines for results-based evaluation: Principles, rationale, planning and managing for evaluations (2013)		To improve the use of evaluation; to harmonize office-wide evaluation practice to support transparency and accountability; and to expand valuation capability in the form of knowledge, skills and tools	The document indicates that the new policy is a collaborative product resulting from contributions of numerous people inside and outside the ILO. The policy focuses on capacity building. It highlights knowledge development, organizational learning and capacity building as priorities of evaluation; encourages stakeholder participation in evaluation and includes rules about engaging them in evaluations; and allocates resources to training and capacity-building activities. It aims to increase knowledge and evaluation capacity in order to facilitate management and ensure mutual accountability.

Document	Context	Purpose of the Policy	Main Focus
New Zealand Aid Program			
Evaluation Policy for the New Zealand Aid Program (2014)	The New Zealand Aid Program is a New Zealand government organization. It delivers New Zealand's official support for developing countries.	To strengthen the role of evaluation in generating evidence for decision-making, learning and accountability	The policy establishes the guiding principles and standards for evaluation; identifies the levels of evaluation; and outlines quality-assurance measures and processes. Even though the evaluations are carried out by external specialists, the policy highlights the importance of building staff capacity and equipping them with the knowledge and tools needed to deliver quality evaluations. The policy identifies roles and responsibilities to respond to evaluation findings and to use them for decision-making.
United Nations Population Fund (UNFPA)			
United Nations Population Fund Evaluation Policy (2013)	UNFPA is an international development organization mandated to promote reproductive health and reproductive rights. It works with 150 countries and territories and partners with governments, civil society organizations and other agencies.	To establish a common institutional basis for the evaluation function and enhance the quality of evaluations; to ensure the independence of the evaluation function; and to guide the conduct of evaluations for organizational learning, management for results and accountability	The policy emphasizes the importance of independent, credible and useful evaluation functions with sufficient resources, and the importance of promoting a culture of evaluation that ensures the active use of evaluation findings and recommendations in developing policy and in improving the functioning of the organization. The policy focuses on adequately supporting and funding evaluation and evaluation-related activities.

Document	Context	Purpose of the Policy	Main Focus
Australian Aid Agency			
Australian Aid Agency Evaluation Policy (2013)	The Australian Aid Agency is an Australian government organization that aims to promote prosperity, reduce poverty and enhance stability.	To set out the minimum expectations for evaluation and performance management for the purpose of improving aid effectiveness	The policy emphasizes accountability through evaluation use. It includes a table that identifies the roles of various stakeholders and what they are accountable for through the use of what. It also includes guiding principles for evaluation and performance management that underscore the importance of usefulness, transparency and clarity.
International Development Research Centre (IDRC)			
IDRC Evaluation and Learning Policy (2017)	IDRC is a key component of Canada's foreign affairs and development efforts. IDRC invests in knowledge, innovation and supports research in developing countries to create real and lasting change. This knowledge can be used as a tool for addressing pressing global challenges.	To describe the organization's approach to evaluation; to explain priorities to various stakeholders; and to strengthen the organization's results-oriented system	The policy stresses the fact that evaluation is framed in utility and that evaluations must have a clearly defined use and must respond to the needs of the user. The guiding principles focus on evaluation utility, quality, collaboration and knowledge sharing. The policy also emphasizes the fact that evaluation processes should develop capacity in evaluative thinking and in evaluation use.

Document	Context	Purpose of the Policy	Main Focus
Spanish Agency for International Development Cooperation (AECID)			
Spanish Cooperation Evaluation Policy (2013)	Part of the Spanish Ministry of Foreign Affairs, AECID is in charge of designing, implementing and managing development cooperation projects and programs—whether directly by using its own resources or indirectly by collaborating with other national and international bodies and non-governmental organizations	To promote transparency, accountability and strategic approaches to evaluation; to optimize resources; and to increase use of evaluation findings and recommendations	The policy establishes a set of principles and guidelines to improve evaluation systems and to help the organization respond to growing complexity in the contexts in which it operates and in the information needs of its various stakeholders, as well as to new demands for evaluation. It supports results-oriented development that entails sophisticated planning and management and the provision of support for organizational culture. The policy focuses on improving evaluation quality, use and capacity through continuous training.
French Development Agency (AFD)			
AFD's Evaluation Policy (2013)	AFD is a public financial institution that implements the international development policy defined by the French government. The organization works to fight poverty and to promote sustainable development in the 108 countries in which it operates, namely by financing, monitoring and	To set out the evaluation principles, standards and approaches to which AFD adheres; to demonstrate, both internally and externally, the emphasis that the institution places on the evaluation of its actions; to inform the various evaluation stakeholders by articulating the rules followed by AFD in designing, conducting and using evaluations; to give clarity to this activity within AFD; and to improve the consistency and	The policy focuses on enhancing the evaluability of the organization's interventions, on improving the allocation of resources to evaluation through more strategic programming, clarifying the roles and responsibilities of the various evaluation stakeholders, and strengthening the quality, utility and use of evaluation. It emphasizes the fact that evaluation is only planned when it is likely to contribute to the achievement of the following objectives: improvement of interventions, increase of knowledge capital and contribution to accountability. The policy

Document	Context	Purpose of the Policy	Main Focus
	supporting over 2,500 development projects.	complementarity of the mechanisms that contribute to the evaluation function or that interact with it	discusses in detail the various types of evaluation use (instrumental, conceptual and symbolic) and explains various actions to promote them. The policy also focuses on capacity building, though the evaluation function is generally outsourced. It emphasizes the fact that a strong in-house capacity is critical in order to be in a position to design and manage complex evaluations, guarantee the quality of the reports produced by the external evaluators and promote the in-house use of evaluations.
Swiss Agency for Development and Cooperation (SDC)			
Swiss Agency for Development and Cooperation Evaluation Policy (2013)	SDC is a component of Switzerland's Federal Department of Foreign Affairs. SDC is responsible for the overall coordination of Switzerland's international development activities and works in cooperation with Eastern European countries as well as with humanitarian aid organizations.	To clarify the status and importance of evaluation in SDC's overall operations; to provide an overview of the evaluation architecture within SDC; and to offer a framework for evaluation.	This policy was developed as part of a required reform of the organization's policy planning and management systems. The organization reshaped and streamlined its procedures and instruments in order to strengthen its results orientation, accountability and institutional learning. The policy was developed to foster an evaluation culture and effective management. It includes 10 guiding principles for the evaluation function and identifies roles and responsibilities. The policy does not discuss the elements of capacity building; rather, it emphasizes independence in evaluation and the importance of outsourcing evaluation in order "to increase transparency."

Document	Context	Purpose of the Policy	Main Focus
Ministry of Foreign Affairs of Japan (MOFA)			
Evaluation Guidelines: Ministry of Foreign Affairs of Japan (2012)	MOFA is the Japanese government organization that's responsible for the country's foreign relations.	To govern the evaluation function; to enhance transparency and accountability; and to guide evaluation management	The document focuses on improving management and maintaining accountability by illustrating in detail the various evaluation duties and by clarifying specific roles and desirable involvement with stakeholders. The document outlines evaluation standards and criteria and the various roles for implementing and managing evaluation systems but does not focus on capacity building or training.
United Nations Institute for Training and Research (UNITAR)			
UNITAR Monitoring and Evaluation Policy Framework (2012)	UNITAR provides training and capacity development activities to assist developing countries.	To develop a more credible and consistent framework for strengthened accountability, organizational learning, quality improvement, and informed decision-making in programming and operations, as well as to contribute to the professionalization of the monitoring and evaluation functions of the organization	The policy discusses and distinguishes between the role of monitoring and the role of evaluation for results-based management. It focuses on using evaluation for strategic planning to achieve results and emphasizes learning from evaluation to improve the quality of the organization's products and services. Taking into account the institution's training-related services, the policy emphasizes the importance of developing evaluation capacity to undertake evaluation activities and to use its findings

Document	Context	Purpose of the Policy	Main Focus
Millennium Development Goals Achievement Fund (MDG-F)			
MDG Achievement Fund Monitoring and Evaluation Policy and Strategy (2012)	MDG is an international development organization that operates in 50 countries, working through more than 27 United Nations agencies in collaboration with citizens, civil society organizations, and local, regional and national governments to target vulnerable groups and tackle multidimensional development challenges	To respond to the needs for accountability, organizational learning and knowledge advancement through evidence-based results reporting; to establish a results-oriented system of monitoring and reporting to inform the MDG Achievement Fund steering committee on the results and progress of the Fund	A principal objective of MDG-F is to replicate and/or scale up successful innovative programs and policies at the country level with the intent of contributing to the achievement of Millennium Development Goals at the global level. This evaluation policy was created to facilitate the achievement of this objective, to fulfil mutual accountability requirements, and to ensure consistency. The policy provides a framework and guidelines for evaluation and emphasizes the fact that the evaluation function is instrumental for accountability and decision-making at the global, national and management levels through the strict application of evaluation guidelines and through follow-ups on evaluation recommendations. The policy recognizes the fact that evaluation is important for increasing the capacity required to support mutual accountability, organizational learning, and the improvement of decision-making and management.
Office for the Coordination of Humanitarian Affairs (OCHA)			
Office for the coordination of Humanitarian Affairs Evaluation Policy (2012)	OCHA is an international development organization responsible for bringing together humanitarian actors to ensure a coherent	To strengthen the institutional framework for the conduct of evaluation activities, to establish a common understanding and approach to the function, and to	The policy includes guiding principles to ensure greater understanding of the evaluation function; to improve the relevance, definition and implementation of key objectives, strategies and policies related to humanitarian

Document	Context	Purpose of the Policy	Main Focus
	response to emergencies. OCHA also ensures that there is a framework within which each actor can contribute to the overall response effort.	provide an overarching framework for the principles, roles and management accountabilities for evaluation within OCHA	coordination; to ensure efficient resource allocation; to improve integration of evaluation as a tool for supporting learning and operational decisions; and to enhance results reporting and accountability at all levels of the organization. The policy focuses on strengthening evaluation capacity throughout the organization by preparing detailed guidelines for conducting evaluation, by conducting staff development and training, and by providing a help-desk function to other units during reviews and assessments.
International Fund for Agricultural Development (IFAD)			
IFAD Evaluation Policy (2011)	IFAD is an international development organization dedicated to eradicating poverty and hunger in rural areas of developing countries.	To describe the purpose of independent evaluation, and the evaluation principles and operational policies to be used by the organization in its independent evaluation work	The policy focuses on independent evaluation and indicates that independence is achieved when evaluation activities are independent from the managers who are responsible for program design, management and implementation. It outlines processes and requirements that are consistent with the purpose of independent evaluation and provides operational and organizational measures to enhance its effectiveness and quality. The policy specifies roles and principles to promote accountability and to foster learning and evaluation use.

Document	Context	Purpose of the Policy	Main Focus
International Federation of Red Cross and Red Crescent Societies (IFRC)			
IFRC Framework for Evaluation (2011)	IFRC is the world's largest humanitarian network. It operates in 190 National Societies through the work of over 17 million volunteers.	To guide how evaluations are planned, managed, conducted and utilized; and to promote reliable, useful and ethical evaluations that contribute to organizational learning, accountability and organization missions to best serve those in need	The framework demonstrates the organization's commitment to improving the importance and utility of evaluation. It focuses on enhancing organizational learning and promoting stakeholders' collaboration to more effectively manage and deliver programs and services. There is a great emphasis on building staff evaluation capacity in order to ensure sufficient experience and expertise to execute evaluation processes that provide credible and useful information for management and decision-making.
United Nations High Commissioner for Refugees (UNHCR)			
UNHCR's evaluation policy (2010)	UNHCR is an international aid organization that is dedicated to saving lives, protecting rights and building a better future for refugees, forcibly displaced communities and stateless people.	To strengthen UNHCR's operational effectiveness and provide managers, staff and partner organizations with useful information about planning, conducting and using evaluations and to enable the organization to engage in effective policymaking, planning, programming and implementation	The organization recognizes the need to develop an organizational culture that is amenable to the development of an effective evaluation function. The evaluation policy is meant to ensure that managers and staff members are fully committed to evaluation and learning, and that the use of evaluation to bring about change is positively encouraged throughout the organization. The policy focuses on enhancing the impact of the evaluation function by pursuing a utilization-focused approach to evaluation and by establishing a steering committee for projects to facilitate this approach

Document	Context	Purpose of the Policy	Main Focus
Organisation for Economic Co-operation and Development (OECD)			
Quality Standards for Development Evaluation (2010)	OECD is an intergovernmental economic organization comprised of 35 member countries that aims to stimulate economic progress and world trade.	To improve the quality of development evaluation processes and products; to support partnerships and collaboration on joint evaluations; to reinforce the contribution of evaluation to improving development outcomes; and to increase the use of evaluation findings	The standards focus on the management and institutional set up of evaluation systems. The document informs evaluation practices and products during the various stages of the evaluation process and in a variety of ways, including assessing the quality of evaluations, informing practice, supporting evaluation training, and fostering learning and use. The standards support evaluation capacity development by improving evaluation knowledge and skills, strengthening evaluation management, stimulating demand for and use of evaluation findings, and supporting an environment of accountability and learning.
Japan International Cooperation Agency (JICA)			
JICA Guidelines for Project Evaluation (2010)	JICA is the Japanese government agency that coordinates official development assistance.	To provide evaluation guidelines that are effective for improving management, ensuring transparency, achieving accountability and fostering learning	The document provides extensive detail about all stages of evaluation and focuses on using evaluation for results-based management and for continuous improvement. It outlines aspects of building and improving mechanisms for utilizing evaluation. The guidelines recognize the importance of stakeholders' involvement for capacity development and indicates that participatory evaluation processes are important for enhancing management, building capacity, facilitating learning and improving accountability.

Document	Context	Purpose of the Policy	Main Focus
Austrian Development Agency			
Guidelines for Project and Program Evaluations (2009)	The Austrian Development Agency is the Austrian government organization responsible for planning, financing and supporting development programs and projects in developing countries.	To support project partners in the process of implementing projects or programs supported by the Austrian Development Agency, in particular in the process of planning, commissioning and managing project and program evaluations; to address the officers of the Austrian Development Agency in charge of contracting and coordinating project or program evaluations; and to explain the evaluation function and roles and to delineate the administrative processes that need to be applied to manage evaluations	The document focuses on evaluation management and quality control. It includes the definitions of the evaluation's purpose, criteria and principles and outlines the required tasks to manage the evaluation to ensure its quality. The document provides templates for an evaluation proposal, for a data-collection planning worksheet, for a management response and for a final evaluation report. The guidelines promote the importance of evaluation utility for decision-making, but it does not include guidelines on building capacity to do evaluation.
Dutch Ministry of Foreign Affairs			
Evaluation policy and guidelines for evaluations (2009)	The Dutch Ministry of Foreign Affairs is the Dutch government organization that is responsible for foreign relations, foreign policy, international development and international trade.	To describe the evaluation policy and the guidelines for evaluations of the Policy and Operations Evaluation Department of the Dutch Ministry of Foreign Affairs, with an emphasis on the placement of the evaluation policy in the context of the Order on Periodic Evaluation and Policy Information (the Netherlands' government-wide framework for the evaluation of	The policy was developed to respond to the growing pressure for ministers to account for the use of public funds and for the results achieved. It explains the various evaluation functions, processes, requirements, roles and responsibilities with a view to increasing transparency and conducting high-quality evaluations for learning and accountability purposes. The policy is intended to strengthen the ministry's capacity to conduct evaluations in partner countries by supporting education

Document	Context	Purpose of the Policy	Main Focus
		government policy)	and training, by participating in joint evaluations, and by promoting the distribution and communication of evaluation results.
United Nations Environment Program (UNEP)			
United Nations Environment Program Evaluation Policy (2009)		To explain the objectives, roles and functions of evaluation within UNEP; define the institutional framework within which it operates; and outline the general processes by which it is operationalized	The policy places an emphasis on results-based management and is meant to enable management to improve programmatic planning, implementing results, monitoring, reporting, and accountability. The policy promotes capacity building, evaluation use and stakeholders' involvement. It includes a table that outlines various evaluation types and the key stakeholders to engage in the evaluation to ensure use and capacity development.

Appendix D: Interview Questions

Part 1: Background Questions

The reason why I asked you to participate in this study is because of your work in the field of evaluation, especially ECB and/or evaluation policy. I am interested in identifying the role of evaluation policy its their influence on the organizational capacity to do and use evaluations. I am also interested in understanding the factors that promote or hinder evaluation-capacity development within the organization.

1. So, first of all, I'd like to ask you about your own definition of:
 - a. Evaluation policy
 - b. Organizational capacity *to do* evaluation?
 - c. Organizational capacity *to use* evaluation?
2. What impact do you think evaluation policy has on organization capacity for evaluation?
 - a. How dose evaluation policy influence evaluation practice?
 - b. How does evaluation policy influence evaluation use?
3. In your opinion, what role does the evaluation policy play in:
 - a. Building the organization's capacity *to do* evaluation?
 - b. Building the organization's capacity *to use* evaluation?
4. Do you think the evaluation policy influences the capacity to do and use evaluation differently? Why?

Part 2: Evaluation Policy in Practice

1. What would you look for in an organization in order to understand the influence of evaluation policy on its evaluation capacity?
2. In your opinion, what strategies can be used to enhance organizational capacity for evaluation?
3. In your opinion, what evidence should organizations collect on the impact of the evaluation policy on evaluation practice and use?

Part 3: Successes and Challenges

1. In your opinion, what factors contribute to benefiting from evaluation policy in building/enhancing organizational capacity for evaluation? (Probe: resources; organizational support; organizational culture; economic, social and political

context)

2. In your opinion, what are the barriers to benefiting from evaluation policy in building/ enhancing organizational capacity for evaluation? (Probe: resources; organizational support; organizational culture; economic, social and political context)
3. Based on your experience, can you think of any concrete steps that can be taken by organizations to address these challenges?

Part 4: Other

1. Did you want to add anything else on the role played by evaluation policy in the organizational capacity to do and use evaluation?

Thank you very much for participating.

Appendix E: Introductory Email (Study 2)

Greetings. I hope all is well with you.

One of my better Ph.D. students, Hind Al Hudib, is doing a three-part study on the relationship of evaluation policy to organizational capacity for evaluation. As you are aware, there is not very much published on evaluation policy, and this is especially the case for empirical studies.

One of Hind's three studies is an interview study with prominent contributors to one or both of these areas. Hind would very much like to interview you because of your prior contributions on evaluation policy.

Hind will follow up with an official invitation in the next few days, along with a letter of informed consent that provides all of the relevant details about the study. I would be very grateful if you would give her request for a single skype interview session serious consideration.

Thanks so much.

J. Bradley Cousins. Ph.D.

Professor Emeritus, Faculty of Education and
Senior Researcher, Centre for Research on Educational and Community Services
University of Ottawa
136 Jean Jacques Lussier
Ottawa, ON, Canada, K1N 6N5
bcousins@uottawa.ca
+1-613-562-5800 ext 4036
+1-613-562-5188 fax

Appendix F: Official Invitation (Study 2)

Dear _____,

I am currently conducting a research project on organizational-evaluation capacity in two regions, namely North America and The Middle East as part of my Ph.D. thesis in organizational studies at the Faculty of Education of the University of Ottawa. Part of my research work involves conducting interviews with evaluation scholars and/or practitioners who have published in the field and who do work in evaluation capacity building. As (describe person's role here), I am seeking your participation in an interview.

My proposed research is designed to explore the role of evaluation policy in the organizational capacity to do and use evaluation in different contexts. The study will provide insights into the role of evaluation policy in shaping evaluation practice and will identify the contextual factors that facilitate or restrain efforts for evaluation capacity building. The results that will emerge from this study will situate research on organizational-evaluation capacity within a broader context of evaluation policy and will provide deeper insight into the dynamics of the organizational capacity for evaluation than is currently available.

The interview should take approximately 45-60 minutes and will be held either in person at a location of your choice, or over the phone.

Please be assured that the information that you will provide will be used only within the context of my Ph.D. thesis work and that your confidentiality will be respected through the use of an alias that will not be linked to your name, electronically or otherwise. With your permission, I would like to record the interviews in order to capture our conversation as accurately as possible. The interview tapes will be transcribed and then kept in a secure location for five years, in accordance with the University of Ottawa's guidelines for ethical research.

I would appreciate if you could let me know by email if you are willing and able to participate in my study. I would like to set up the first interview over the next two months and therefore would need to find out your preference for a date and time.

I would be happy to answer any questions that you might have regarding the research project or the interview prior to our first meeting, or when we meet.

Thank you very much for taking the time to consider this request. I look forward to hearing your insights on this interesting topic.

Sincerely,

Hind Al Hudib, Ph.D. Candidate
Faculty of Education
University of Ottawa
Canada

Appendix G: Informed Consent

Informed Consent Letter of Participation in the Interview (Study 2)

Title of the study: The Role of Evaluation Policy in Organizational Capacity to Do and Use Evaluation

Invitation to participate: I am invited to participate in the research project conducted by Ms. Hind Al Hudib, a Ph.D. candidate at the Faculty of Education of the University of Ottawa.

Purpose of the study: The purpose of the study is to develop a deep understanding of the role evaluation policy plays in shaping organizational capacity to *do* evaluation and to *use* evaluation.

Participation: My participation involves an in-person interview, taking approximately 45-60 minutes with the researcher on the role played by evaluation policy in the organizational capacity to do and use evaluation. The interview will be conducted at a public location of my choice or over the phone.

Risk: I understand that, since these activities deal with my professional opinion, I may feel uncomfortable with some of the questions asked during the interview or the focus group. I have received assurances from the researchers that every effort will be made to minimize these risks. Moreover, I am free to withdraw from the research project at any time, to refuse to go on, or to refuse to answer particular questions.

Benefits: My participation in this study may lead to personal benefits derived from a reflection on my own evaluation experience and practice and will contribute to the advancement, development and transmission of knowledge in the field of evaluation.

Confidentiality and anonymity: I understand that the information that I will provide will be used only within the context of the researcher's Ph.D. thesis project and that my confidentiality will be respected through the use of an alias that will not be linked to my name, electronically or otherwise. I also understand that my confidentiality cannot be guaranteed in the context of the focus group, given the presence of other participants. I understand that both the interview and the focus group will be audio recorded; the transcripts produced from the recordings will only be available to the principal researcher and the thesis supervisor. I will be given access to my password-protected interview transcripts via email.

Conservation of data: The data will be managed only in electronic form. Audio recordings and transcripts of interviews and focus groups will be stored in a password protected project folder with access only by the supervisor and the researcher and will be destroyed after five years after the date of the focus group.

Voluntary Participation: I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and/or refuse to answer any questions, without suffering any negative consequences. If I choose to withdraw from the study, I can also withdraw my individual interview data; however, given the interdependent nature of focus group transcripts, I understand that individual data from these sessions cannot be withdrawn.

Acceptance: I, _____, agree to participate in the above research study conducted by Ms. Hind Al Hudib of the Faculty of Education, University of Ottawa, which research is under the supervision of Professor Brad Cousins.

I understand that, by signing below, I am providing my consent to participate in the study.

If I have any questions about the study, I may contact the researcher or his supervisor.

If I have any questions regarding the ethical conduct of this study, I may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5 Tel.: +1-613-562-5387

Email: ethics@uottawa.ca

There are two copies of the consent form, one of which is mine to keep.

Signature of participant: _____ Date: _____

Signature of researcher: _____ Date: _____

Principal Researcher:

Hind A. Al Hudib, Ph.D. Candidate
Faculty of Education, University of Ottawa
Center for Research on Educational and Community Services (CRECS)
Vanier Room 5002
136 Jean Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5

Thesis Supervisor:

Professor Brad Cousins, Ph.D.
Faculty of Education, University of Ottawa
Center for Research on Educational and Community Services (CRECS)
Vanier Room 5002
136 Jean Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5
Tel.: (613) 562-5800 extension 4036
Email: (bcousins@uottawa.ca)

Appendix H: Research Background (Study 3)

Evaluation Policy and Organizational Capacity for Evaluation

By Hind Al Hudib, Ph.D. Candidate, University of Ottawa.

This short paper provides an overview of the research and a summary of preliminary findings to be read in preparation for the upcoming focus group. The purpose of the focus group is to discuss the findings in terms of their implications for organizational evaluation practice. I also am interested in ascertaining the extent to which the findings resonate with the experiences of evaluators and organization members working in Turkey and the way evaluation policy influences organizational evaluation capacity and evaluation practice in developing- and transitional- country contexts.

The overarching objective of the study is to gain a deeper understanding of the role played by evaluation policy in organizational capacity for evaluation by exploring the following questions:

- What is the role of evaluation policy in the organizational capacity to do and use evaluation?
- How does evaluation policy enhance (or impede) the integration of evaluation into organizational culture?
- What are the contextual facilitators and barriers that influence the role of evaluation policy in enhancing the integration of evaluation into organizational culture?

In this document, I present for discussion the findings of (i) an integrated review of 50 evaluation policies from 35 organizations and bi-lateral institutions and (ii) a set of 18 interviews conducted with leading evaluation scholars and practitioners from Canada, the United States and Europe who have published substantial work in relation to evaluation capacity building (ECB) and evaluation policy. The findings of the study are critical for developing the basis of a more focused and in-depth understanding of evaluation policy and its role in building organizational evaluation capacity.

Findings for Discussion

I found that the role of evaluation policy in leveraging organizational evaluation capacity was quite limited, although this was not always the case. Most interview participants agree that in and of itself, evaluation policy has little relationship with an organization's capacity to implement evaluation or to use it. The extent to which evaluation policy will positively affect capacity will depend on a number of conditions or moderating variables. From my analyses I identified 12 such variables: 1) context (political, social, cultural, economic); 2) characteristics of evaluation policy (content, type); 3) supportive evaluation system; 4) resources; 5) leadership; 6) evaluative culture; 7) evaluator role; 8) interpersonal relations; 9) integration of evaluation into decision-

making processes; 10) purpose of evaluation function; 11) stakeholder engagement; and 12) learning.

I found these conditions or variables that moderate evaluation policy influences on capacity to be operating at one of three levels, as is set out in Figure 1. At the *macro* level social and political contextual conditions and features or characteristics of often externally imposed policies enhance or impede policy's influence on organizational capacity to do and use evaluation at the organizational—or *meso*—level as well as at the stakeholder—or *micro*—level. The data reveal that the macro-level conditions have different moderating influences on the capacity to do and the capacity to use evaluation at the organizational level and at the stakeholder level.

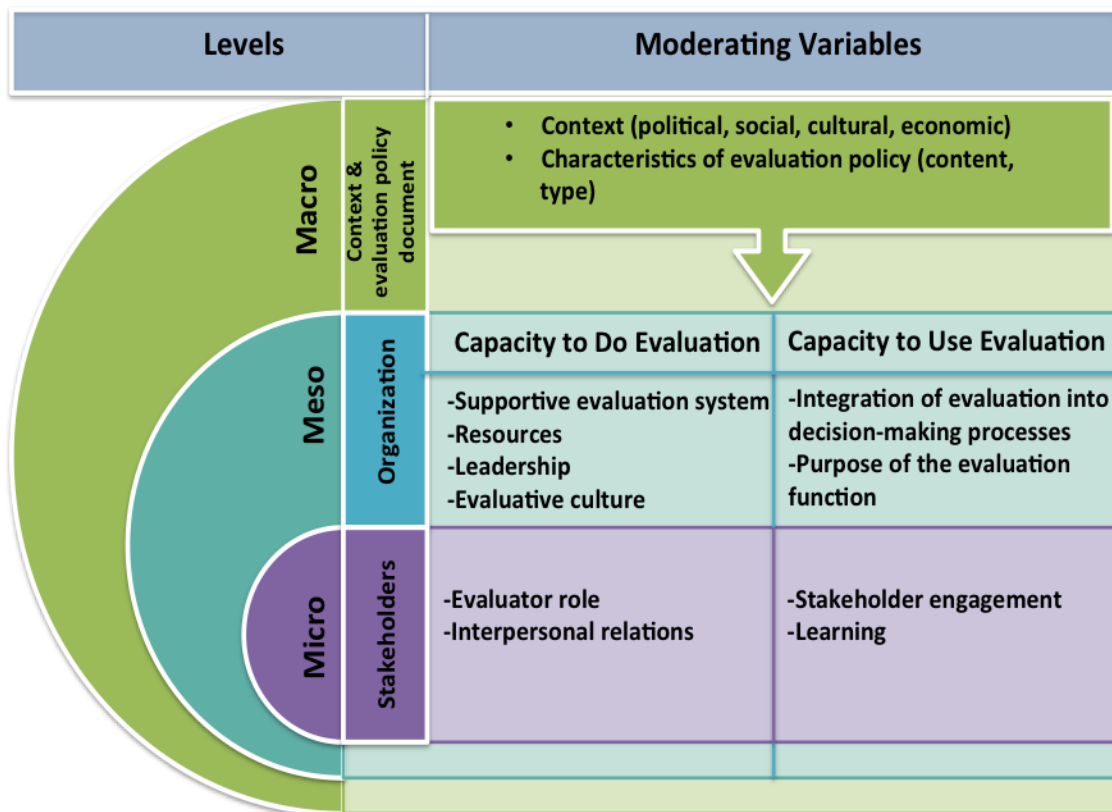


Figure 1: Moderating variables that explain the relationship between evaluation policy and organizational capacity for evaluation

As shown in Figure 1, evaluation policy's impact on the capacity to do evaluation at the organizational level is affected by the availability of the evaluation budget, the commitment of the leadership, the presence of an evaluative culture that accepts and values evaluation, and the plan for and processes of implementing the evaluation policy. These factors have a direct influence on the role of the evaluator and on interpersonal relations among stakeholders as individual-level variables moderating policy effects on the capacity to do evaluation.

On the other hand, evaluation policy's influence on the capacity to use evaluation at the organizational level is influenced by the integration of evaluation into decision-making processes, a key condition that helps to leverage action. The other moderating condition of note is organization member accountability to the evaluation policy. At the individual level, stakeholders' engagement in the evaluation process and their learning from evaluation activities are the two conditions that were identified as moderating policy influence on their capacity to use evaluation.

The data also reveal that evaluation policy plays an important role in developing better and more effective organizational capacity for evaluation if it clearly defines ECB processes and activities for meeting immediate and long-term organizational needs and priorities. An effective evaluation policy allows governments, organizations, funding agencies, evaluators and stakeholders to have a common understanding about the:

- Definition of the capacity to do and use evaluation (i.e., what constitutes capacity building, what is the organizational capacity to do and use evaluation);
- Objectives of organizational capacity for evaluation (i.e., what capacity building should seek to achieve within the organization);
- Key conditions or variables that moderate policy's influence on the capacity to do and use evaluation at the organizational and stakeholder levels (i.e., the conditions identified in Figure 1); and
- Systems and activities that are available to build and support the organizational capacity for evaluation at the organizational and individual levels (e.g., training and on-the-job training, policy dialogues).

I present these findings for discussion with members of the evaluation community in Turkey about the role played by evaluation policy in the organizational capacity for evaluation.

Focus Group Guiding Questions

Participants will be provided with a short report about the findings from the interviews in study 2 drafted by the researcher and will be asked to provide input and feedback.

I am interested in hearing about the members' views on questions such as:

- To what extent do organization evaluation policies exist in the region?
- Does evaluation policy play a role in fostering organizational capacity for evaluation? To do evaluation? Or to use it?
- If yes, what conditions or variables need to be in place?

Appendix I: Informed Consent (Study 3)

To: Members of evaluation community in Turkey/ Jordan

Re: The Role of Evaluation Policy in Organizational Capacity to Do and Use Evaluation

I am writing to invite you to participate in a research project conducted by myself, Hind Al Hudib, and my thesis supervisor, Professor Brad Cousins, of the Faculty of Education, University of Ottawa, Canada.

What is the purpose of the study? The purpose of this study is to develop a deep understanding of the role played by evaluation policy in shaping organizational capacity to *do* evaluation and to *use* evaluation.

What will my participation involve? We are asking for your participation in a **focus group** meeting of approximately 60 to 90 minutes to discuss the role played by evaluation policy in organizational capacity to do and use evaluation. The meeting will be held in Ankara/ Amman (Hotel) on agreed time. In advance of the meeting we ask that you read a short background document on the topic.

How will the information be used? The information that you provide will be used only within the context of the researcher's Ph.D. thesis project. Our commitment is to treat the information as confidential: all names of participating individuals and organizations will be withheld from any reports arising from the research. Since there will be other participants in the focus group we cannot guarantee confidentiality although we will ask everyone present to respect this wish. The focus group will be audio-recorded; transcripts will only be available to the principal researcher and to the thesis supervisor.

What are the risks of my participation? There are no appreciable risks associated with participation. Since these activities deal with your professional opinion there is a modest risk that you may feel uncomfortable with some of the questions asked during the focus group. It is important to note that you are free to withdraw from the research project at any time, to refuse to continue or to refuse to answer particular questions.

What are the benefits of participation? Your participation in this study may lead to personal benefits derived from reflection on your own evaluation experience and practice and will contribute to the advancement, development and transmission of knowledge in the field of evaluation. Once the study is completed, a summary of the findings will be circulated to you.

How will the information be conserved? The data will be managed only in electronic form. The audio recordings and transcripts of the interviews and focus groups will be stored in a password-protected project folder, with access only being available to the supervisor and the researcher, and will be destroyed five years after the date of the focus group.

What does voluntary participation mean? You are under no obligation to participate and you can choose to withdraw from the study at any time and/or refuse to answer any questions without suffering any negative consequences. If you choose to withdraw from the study, it is important to understand that individual data from the focus group cannot be withdrawn given the interdependent nature of the transcripts.

Acceptance:

I, _____ (name), agree to participate in the above-described research study conducted by Ms. Hind Al Hudib of the Faculty of Education, University of Ottawa, under the supervision of Professor Brad Cousins.

I understand that, by signing hereunder, I am providing my consent to participate in the study.

If I have any questions about the study, I may contact the researcher or her supervisor.

If I have any questions about the ethical conduct of this study, I may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5 Tel.: +1-613-562-5387

Email: ethics@uottawa.ca

There are two copies of the consent form, one of which is mine to keep.

Signature of the participant: _____ Date: _____

Signature of researcher: _____ Date: _____

Principal Researcher:

Hind A. Al Hudib, Ph.D. Candidate
Faculty of Education, University of Ottawa
Centre for Research on Educational and Community Services (CRECS)
Vanier Room 5002
136 Jean Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5

Thesis Supervisor:

Professor Brad Cousins, Ph.D.
Faculty of Education, University of Ottawa
Centre for Research on Educational and Community Services (CRECS)
Vanier Room 5002
136 Jean Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5
Tel.: 613-562-5800, extension 4036
Email: bcousins@uottawa.ca