



uOttawa

L'Université canadienne
Canada's university

**FACULTÉ DES ÉTUDES SUPÉRIEURES
ET POSTDOCTORALES**



**FACULTY OF GRADUATE AND
POSTDOCTORAL STUDIES**

Isabelle Bourgeois

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Education)

GRADE / DEGREE

Faculty of Education

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Understanding the Dimensions of Organizational Evaluation Capacity

TITRE DE LA THÈSE / TITLE OF THESIS

Bradley Cousins

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Swee Goh

Marielle Simon

John Ross

David Smith

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Understanding the Dimensions of Organizational Evaluation Capacity

by

Isabelle Bourgeois

Faculty of Education

Faculty of Graduate and Postdoctoral Studies

A thesis submitted in conformity with the requirements
of the degree of Doctor of Philosophy
University of Ottawa

© Isabelle Bourgeois 2008



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-48389-3
Our file *Notre référence*
ISBN: 978-0-494-48389-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

The question of evaluation capacity in governmental organizations has been studied empirically in only a limited fashion in recent years. The actual dimensions of evaluation capacity, or what evaluation capacity might look like in an organization, have not previously been identified systematically in the literature. This study sought to identify the key dimensions of evaluation capacity in Canadian federal government organizations as well as the manageable steps required to move from low to high capacity for each of these.

The data collection approach was articulated around three phases. The first phase sought to identify the key dimensions of evaluation capacity in Canadian federal government organizations and characterize minimal and exemplary performance on each dimension through the development of a framework. This was achieved through a literature review and semi-structured key informant interviews. The second phase focused on refining and clarifying these dimensions by using key informant interviews. The third and final phase sought to validate the revised framework. This was achieved through its practical application in four federal government organizations and semi-structured interviews with those involved in applying the framework.

The study concluded that evaluation capacity in Canadian federal government departments and agencies can be described through six main dimensions, each one broken down into further sub-dimensions. The capacity of these organizations on each of these dimensions and sub-dimensions can be assessed using four levels: low, developing, intermediate, and exemplary. The study found that organizations vary in terms of their capacity from one dimension to the next, and indeed, from one sub-dimension to the next. This study makes an important contribution to the body of knowledge on evaluation capacity; although much has been published

in the evaluation literature on evaluation capacity building, the actual characteristics and attributes of evaluation capacity itself have not before been defined and described based on empirical data. Continuing research may focus on expanding the scope of the framework, while the main practical implication of this study is the use of the framework as an instrument of evaluation capacity for government organizations.

DEDICATION

This thesis is dedicated first and foremost to my husband Shawn. His unwavering encouragement and continued assistance (as driver, childcare provider, technical support specialist, cheerleader, to name but a few) were invaluable throughout all of my graduate studies. I also dedicate this work to my parents, who have been a great source of support and inspiration throughout my life. I could not imagine what life would be like without them. Finally, I want to dedicate this thesis to my son Jérémie, whose arrival prompted me to work even harder to finish my studies. I hope to be a good role model to him, now and in the years to come.

ACKNOWLEDGEMENTS

Many people have contributed to this project and I wish to extend my gratitude to all of them. I would first like to thank my advisor, Dr. Brad Cousins, for his continued encouragement and for sharing his knowledge and expertise with me throughout my doctoral studies. His availability and generosity with his time, despite an incredibly busy schedule, were much appreciated and made it possible for me to keep my project on track. His views on evaluation capacity have greatly influenced the direction of my work, both as a researcher and as an evaluation practitioner.

I also wish to thank the members of my thesis committee, Drs. Swee Goh, Marielle Simon, and Martin Barlosky, for their thoughtful comments and valuable suggestions at each step of the research process. I greatly benefited from the diversity of their views on evaluation and evaluation capacity. I am also grateful to Dr. John Ross for his comments as external examiner, and to Dr. David Smith, who graciously agreed to step in as internal examiner for the defense.

I was very fortunate to have the support of fellow Ph.D.s-in-Progress (PiPs) over the last four years. Sharing this journey with a group of caring, intelligent and fun people was invaluable. Thanks in particular to Claire, Nicola, Darene, Robin, Alessandra, and Catherine. Thanks also to my friend and co-conspirator Courtney. This experience would not have been the same without you.

Heartfelt thanks go to a number of others without whose help this thesis would not have been possible. I am grateful to the fifteen individuals who so generously participated in the study, to Susan Morris and Steve Montague, for being such wonderful mentors early in my

evaluation career, and to my new colleagues at the National Research Council for their encouragement and support.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
Statement of the Problem.....	1
Conceptual Distinction Between Evaluation Capacity and ECB	3
Overview of the Study	3
<i>Context for the Study</i>	4
<i>Summary of Methodology</i>	6
Overview of the Thesis	7
CHAPTER 2: REVIEW OF THE LITERATURE.....	9
Rationale for Evaluation Capacity Building.....	10
Evaluation Use.....	11
Organizational Learning	15
Organizational Factors Contributing to the Success of ECB.....	18
<i>External Environment</i>	18
<i>Organizational Structure</i>	20
<i>Organizational Culture</i>	23
<i>Organizational Leadership</i>	24
Planned Organizational Change.....	26
<i>Barriers and Facilitators to Organizational Change</i>	27
<i>Models of Organizational Change</i>	28
<i>Stages of Change</i>	30
Stages of Evaluation Capacity Building	32
Summary: A Conceptual Model of Evaluation Capacity Building	34
Evaluation in the Canadian Federal Government	35
Alternatives to Evaluation Capacity Building	37
<i>Mainstreaming</i>	37
<i>Insourcing</i>	38
Theoretical Critique of ECB.....	39
<i>Governmentality and Evaluation</i>	40
<i>Disciplinary Power and the Evaluator</i>	41
Profiling Organizational Growth in Evaluation Capacity.....	42
Challenges to the Implementation and Study of Evaluation Capacity	44
Research Questions.....	46
CHAPTER 3 – METHODS	48
Overview of Research Design	48
<i>Basis for Selected Approach</i>	48
<i>Epistemological Assumptions</i>	51
<i>Role of Researcher</i>	52
Phase 1: Identification of Key Dimensions of Evaluation Capacity (Divergent Phase).....	54
<i>In-depth Literature Review: Procedures and Data Analysis</i>	55
<i>Semi-Structured Interviews: Participants</i>	56
<i>Semi-Structured Interviews: Procedures</i>	59

<i>Semi-Structured Interviews: Data Collection</i>	59
<i>Semi-Structured Interviews: Data Analysis and Reporting</i>	61
Phase 2: Review and Feedback on Draft Framework (Convergent Phase)	62
<i>Procedures</i>	63
<i>Data Collection</i>	63
<i>Data Analysis and Reporting</i>	65
Phase 3: Triangulation of Findings Included in the Framework.....	66
<i>Organizations Selected for the Validation Exercise</i>	66
<i>Participants</i>	68
<i>Data Collection</i>	72
<i>Data Analysis and Reporting</i>	73
CHAPTER 4 – FINDINGS FROM PHASE 1	75
In-Depth Review of the Literature.....	75
<i>Description of Dimensions Included in Preliminary Framework</i>	78
Interview Process and Findings	82
<i>Differences Between Literature and Interview Data</i>	83
Summary of Revisions for Phase 1	90
Markers of Organizational Evaluation Capacity.....	97
Capacity of Specific Departments and Agencies.....	98
<i>High Capacity Organizations</i>	99
<i>Intermediate Capacity Organizations</i>	99
<i>Typical or Developing Capacity Organizations</i>	100
Selection of Organizations for Phase 3 of the Study	101
CHAPTER 5 – FINDINGS FROM PHASE 2	102
General Changes Made to Framework	112
<i>Front Page Summary</i>	112
<i>Names of Levels</i>	112
<i>Number of Levels</i>	114
<i>Location of Sub-Dimensions</i>	114
Changes to Specific Dimensions	115
<i>Human Resources (dimension 1)</i>	115
<i>Organizational Resources (dimension 2)</i>	118
<i>Evaluation Planning and Activities (dimension 3)</i>	120
<i>Organizational Characteristics (dimension 4)</i>	122
<i>Integration with Organizational Decision-Making (dimension 5)</i>	123
<i>Learning Benefits (dimension 6)</i>	123
Other Comments on Framework.....	124
Summary of Findings for Phase 2.....	124
CHAPTER 6 – FINDINGS FROM PHASE 3	126
General Comments on Framework	126
Comments on Sub-Dimensions	139
<i>Sub-dimension 1.1: Staffing</i>	139
<i>Sub-dimension 1.2: Evaluation Logic and Technical Skills</i>	140
<i>Sub-dimension 1.3: Communications and Interpersonal Skills</i>	141

<i>Sub-dimension 1.4: Professional Development</i>	141
<i>Sub-dimension 1.5: Leadership</i>	142
<i>Sub-dimension 2.1: Budget</i>	142
<i>Sub-dimension 2.2: Ongoing Data Collection</i>	143
<i>Sub-dimension 2.3: Organizational Infrastructure</i>	143
<i>Sub-dimension 3.1: Evaluation Planning and Activities</i>	144
<i>Sub-dimension 3.2: Use of Consultants</i>	145
<i>Sub-dimension 3.3: Information Sharing</i>	145
<i>Sub-dimension 3.4: External Supports</i>	146
<i>Sub-dimension 3.5: Organizational Linkages</i>	146
<i>Sub-dimension 4.1: Involvement in Evaluation</i>	147
<i>Sub-dimension 4.2: Results-Management Orientation</i>	147
<i>Sub-dimension 5.1: Management Processes</i>	148
<i>Sub-dimension 5.2: Decision Support</i>	148
<i>Sub-dimensions 6.1: and 6.2 Instrumental/Conceptual Use and Process Use</i>	148
Summary of Findings on Framework.....	148
Organizational Capacity Findings.....	149
<i>Organization 1 – Exemplary Level</i>	153
<i>Organization 2 – Exemplary Level</i>	159
<i>Organization 3 – Intermediate Level</i>	165
<i>Organization 4 – Low Level</i>	170
Observations on Organizational Capacity Findings.....	175
CHAPTER 7 – DISCUSSION	177
Motivation for the Study.....	177
Overview of Study.....	177
Summary of Findings by Research Question.....	179
<i>Dimensions of Evaluation Capacity in Government Organizations</i>	179
<i>Capacity Levels – Minimal to Exemplary Performance</i>	183
Relevant Themes Emerging from the Literature.....	185
<i>Evaluation Use</i>	186
<i>Organizational Factors Contributing to the Success of ECB</i>	188
<i>Stages of Evaluation Capacity Building</i>	191
<i>Implications for Conceptual Model of Evaluation Capacity</i>	193
<i>Other Important Findings</i>	194
<i>Implications for Theory</i>	195
Limitations of the Study.....	206
Continuing Research.....	209
<i>Scope of Framework</i>	209
<i>Structure of Framework</i>	211
<i>Alternative Research Methods</i>	211
<i>Other Development Models</i>	212
<i>Organizational Change</i>	213
Implications for Practice and Policy.....	213
<i>Capacity to Do Evaluation</i>	214
<i>Capacity to Use Evaluation</i>	215
<i>Framework as Instrument of Evaluation Capacity</i>	217

Conclusion	218
REFERENCES	223
APPENDIX A	233
Phase 1 - Introductory Email	233
APPENDIX B	234
Phase 1 - Letter of Informed Consent	234
APPENDIX C	235
Phase 1 - Interview Protocol	235
APPENDIX D	237
Phase 2 - Interview Protocol	237
APPENDIX E	238
Phase 3 - Introductory Email	238
APPENDIX F	239
Phase 3 - Letter of Informed Consent	239
APPENDIX G	240
Phase 3 - Interview Protocol	240
APPENDIX H	241
Phase 3 - Letter of Organizational Consent	241

LIST OF TABLES

- Table 1: Participants by Phase of Study (page 54)
Table 2: Preliminary Framework Based on Literature Review (page 77)
Table 3: Phase 1 Framework (page 91)
Table 4: Phase 2 Framework (page 104)
Table 5: Phase 3 Framework (page 129)
Table 6: Summary of Findings by Organization (page 151)

LIST OF FIGURES

- Figure 1: Conceptual Model of Evaluation Capacity Building (page 34)
- Figure 2: Summary of Methodology (page 51)
- Figure 3: Summary of Phase 2 Framework (page 103)
- Figure 4: Summary of Phase 3 Framework (page 128)
- Figure 5: Revised Conceptual Model of Evaluation Capacity Building (page 194)

LIST OF APPENDICES

Appendix A	Phase 1 – Introductory Email
Appendix B	Phase 1 – Letter of Informed Consent
Appendix C	Phase 1 – Interview Protocol
Appendix D	Phase 2 – Interview Protocol
Appendix E	Phase 3 – Introductory Email
Appendix F	Phase 3 – Letter of Informed Consent
Appendix G	Phase 3 – Interview Protocol
Appendix H	Phase 3 – Letter of Organizational Consent

LIST OF ACRONYMS AND ABBREVIATIONS

ACS	American Cancer Society
ADM	Assistant Deputy Minister
CDC	Centers for Disease Control
DM	Deputy Minister
DPR	Departmental Performance Report
ECB	Evaluation Capacity Building
HR	Human Resources
MAF	Management Accountability Framework
MC	Memorandum to Cabinet
OL	Organizational Learning
PAA	Program Activity Architecture
P-PE	Practical Participatory Evaluation
RMAF	Results-based Management and Accountability Framework
RPP	Report on Plans and Priorities
TBS	Treasury Board Secretariat
TTM	Transtheoretical Model

CHAPTER 1: Introduction

Statement of the Problem

Service organizations around the world have been moving from rules-focused management practices to more citizen-centred public administration through various initiatives. One initiative undertaken by the government of Canada to move in this direction, Results for Canadians (2000), proposes the use of Results-based Management (RBM) as the mechanism through which this can be achieved. RBM focuses on the measurement of progress towards sought after results, or outcomes, and modifying program design and operations based on the measures obtained. This approach is meant to improve the accountability of public institutions while providing useful information to those managing the programs in question (Jorjani, 1998).

Managing for results requires a comprehensive system of performance measurement and program evaluation. Taken together, these evaluative activities have the potential to provide a clear picture of a program's progress towards pre-determined outcomes. Recent federal government policies place the responsibility for performance measurement in the hands of program managers, who are also required to actively contribute to program evaluation efforts usually undertaken by corporate evaluation officers. However, guidance and resources provided from the Treasury Board Secretariat have been developed for trained evaluators rather than program managers and staff less familiar with program measurement and evaluation techniques. This limits the applicability and usefulness of performance measurement and evaluation activities because they are typically conducted by evaluators and somewhat removed from the realities of everyday program administration and delivery.

The implementation of evaluation capacity building (ECB) initiatives within federal government organizations, therefore, offers a potential bridge between the technical expertise required to conduct evaluative activities and the substantive knowledge of program managers and staff. ECB refers to the changes undertaken by organizations to integrate evaluation practice and use at all levels (Boyle, Lemaire & Rist, 1999; Cousins, Goh, Clark & Lee, 2004; Sanders, 2002; Stockdill, Baizerman & Compton, 2002). ECB is defined by Stockdill et al. (2002) as:

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

The development and strengthening of an organization's evaluation capacity has the potential to lead to significant organizational improvement in two ways. The first is through the participatory approach – an approach involving the collaborative actions of trained evaluation personnel working with program and organization personnel not trained in evaluation – which has been shown to lead to increased use of the evaluation findings and process (Cousins & Leithwood, 1993; Johnson, 1998; Patton, 1999). Second, ECB fosters organizational learning by bringing together stakeholders and fostering “communities of practice” involved in experimentation and learning (Kruse, 2003; Sanders, 2002; Wenger, 2004).

To date, the extent to which large governmental organizations have developed their evaluation capacity has not been the focus of much empirical research. In fact, the actual dimensions of evaluation capacity as they could be observed in organizations have not yet been identified in the literature. Knowledge on these dimensions may help organizations identify what efforts are required to attain a greater level of evaluation capacity. The thesis identifies the key

dimensions of evaluation capacity as they appear in organizations. These dimensions were operationalized through a framework based on the Innovation Profiles method developed by Leithwood and Montgomery (1987).

Conceptual Distinction Between Evaluation Capacity and ECB

An important feature of this study is the distinction made between evaluation capacity building as defined above and an organization's actual evaluation capacity. Evaluation capacity can be generally defined as an organization's visible, enacted evaluation practices and processes, whereas ECB is the process by which an organization develops its understanding and ability to undertake these practices and processes. In reality, these two concepts are closely intertwined, and, as will be demonstrated in the literature review, little attention has been paid up until now to the conceptual differences between the two. The study presented here focuses specifically on the construct of evaluation capacity, temporarily leaving behind the notion of evaluation capacity building, in order to describe evaluation capacity as it may appear in organizations. One way in which the distinction between ECB and evaluation capacity can be conceptualized is through an analogy based on logic models. In this conceptualization, ECB would represent the activities and outputs of a program, whereas evaluation capacity would represent its immediate and intermediate outcomes.

Overview of the Study

The study addresses a lack of empirically-derived knowledge on the key dimensions of evaluation capacity by examining the following questions:

1. What are the essential dimensions of evaluation capacity in Canadian federal government organizations?

2. How are minimal and exemplary performance on each of these dimensions characterized?
3. What are the manageable steps required to move from minimal to exemplary performance?

In essence, the objective of the research was to identify the key dimensions of evaluation capacity in Canadian federal government organizations as well as the manageable steps required to move from low to high capacity for each of these. The dimensions of evaluation capacity and the steps identified for each dimension were captured in the form of a key component profile framework and validated in government organizations.

Context for the Study

The Canadian federal government is comprised of 113 departments and agencies (known as the core public service) as well as a variable number of other organizations serving various purposes (crown corporations, special operating agencies, etc). Of these 113 organizations making up the core public service, 87 are required to report on results through a formal evaluation process by the Treasury Board Evaluation Policy (2001). The Evaluation Policy applies to all organizations considered to be departments within the meaning of section 2 of the Financial Administration Act. The most recent information available on the evaluation function in the government of Canada, published in 2004 (Government of Canada, 2004), estimates that of the 87 organizations required to evaluate their programs, policies and initiatives, 42 report an in-house evaluation function, while the 45 others have no permanent evaluation capacity. These estimates also place the number of evaluation professionals working in government to about 300 individuals.

The federal government publishes approximately 250 evaluations per year, in addition to hundreds of other reports relating to evaluation or performance measurement, such as Results-based Management and Accountability Frameworks (RMAFs).

The current Evaluation Policy, released in 2001, requires departments and agencies to evaluate their programs according to three main issues: relevance, success, and cost-effectiveness. Individual organizations are free to determine, through a transparent and consultative evaluation planning process, the programs to be evaluated and the scope of the studies to be undertaken. The spirit of the 2001 Evaluation Policy speaks largely to a concern with results-based management, as described in the first section of this chapter; evaluation is therefore seen as a learning mechanism, and the processes put in place to identify evaluands, the issues to be addressed and the extent to which program personnel are involved focus primarily on organizational learning. However, a new evaluation policy, set to be implemented in 2008, is thought to have a narrower focus on accountability rather than learning, with new requirements for complete coverage of all programs within any given organization every five years, with no concrete promises of new resources to be allocated to departments to fulfill this new mandate. In addition to this, the proposed new policy clearly identifies the client of evaluation as being the deputy minister rather than the program manager, which further supports the summative viewpoint that is to be taken by evaluators. No longer will evaluators be required to help their program counterparts develop their own evaluation and performance measurement skills; instead, the emphasis will be placed on the evaluators' neutrality and objectivity. Although this new policy is not yet in place, informal discussions held at various meetings about these new elements have cast some doubt on the capacity of departments and agencies to meet these requirements. As organizations prepare for the launch of the new policy, there is a clear shift

occurring in the discourse of evaluators feeling the pressure to change from a learning role to an accountability role. As can be expected, this will have a significant impact on the evaluation capacity of federal government departments and agencies. This change in mentality is also apparent in the comments of those participating in this study. Further discussion of this matter will be interspersed throughout the results and discussion sections of the thesis.

Summary of Methodology

The data collection approach was articulated around three phases. The first phase sought to identify the key dimensions of evaluation capacity in Canadian federal government organizations and characterize minimal and exemplary performance on each dimension through the development of a framework. This was achieved through a literature review of mostly anecdotal reports of evaluation capacity, as well as semi-structured key informant interviews. An important aspect of the literature review involved moving beyond descriptions of capacity building initiatives undertaken in various organizations and focusing instead on the definitions and features of evaluation capacity itself and how it appears in organizations. This distinction is not often made in the literature; as mentioned previously, evaluation capacity is seldom defined in operational terms. The second phase focused on refining the framework based on interview data collected with the same expert participants as the first phase. Participants were asked to comment on both the contents and wording of the dimensions included in the framework. The third and final phase of data collection sought to validate the revised framework in an organizational setting. The validation exercise was based on whether the dimensions, as outlined in the framework that was developed as an outcome of the study, enabled organizations to better understand their current level of evaluation capacity and provided them with useful information on potential next steps. This was achieved through the use of the framework in four federal

government organizations and semi-structured interviews with those involved in applying the framework.

Overview of the Thesis

The first few chapters of the thesis introduce the study reported here by way of a literature review (Chapter 2) and a detailed description of the methodology used (Chapter 3). The literature review focuses on a number of elements related to evaluation capacity building, such as the factors thought to influence ECB in organizations and the stages of ECB. An analysis of organizational change is also presented in the literature review to situate ECB in the change process. This focus on ECB, rather than on the more specific dimensions of evaluation capacity, is mainly due to the fact that almost all of the publications dealing with evaluation capacity focus on its development rather than its features or effects on organizations. Wherever available, sources focusing on evaluation capacity itself are outlined and summarized. A brief outline of the history of evaluation in the Canadian government is also provided in the literature review in order to set the stage for the study. In addition to this, some alternatives to ECB as well as a critique of evaluation in general are presented to provide some counterbalance to the discussion. The critique of evaluation is based on Foucault's discussion of "governmentality". The methods chapter presents the details of the procedures followed during all three phases of data collection.

Chapters 4, 5, and 6 of the thesis describe the findings obtained in the three phases of data collection. Phase I findings are presented in Chapter 4 and provide the first draft of the framework developed to summarize the in-depth literature review and interview findings. Chapter 5 presents a revised version of the framework, based on the interview process undertaken in the second phase of the study. Finally, Chapter 6 provides a final version of the framework, based on the validation exercise completed in the third phase of the study.

Chapter 7 of the thesis presents a detailed discussion of the findings, the implications that these may have on research and practice, as well as recommendations and a final conclusion.

Chapter 2: Review of the Literature

The purpose of this literature review is to highlight some of the key theories, models and empirical findings on evaluation capacity building in service organizations as well as its major components. The literature review provides a summary of the current knowledge base on ECB in order to set the stage for the study. As described in the previous chapter, the literature review focuses primarily on evaluation capacity building, rather than the concept of evaluation capacity itself, due to the lack of information on the latter in the literature. Because evaluation capacity building is such an important part of an organization's capacity to do and use evaluation, the literature summarized and critiqued here remains relevant to this study. It should be noted, however, that even though the two concepts are separated for clarity, they are closely intertwined in practice.

The review starts by outlining the rationale for ECB and its two key consequences: evaluation use and organizational learning. The organizational factors critical to ECB are then presented and are followed by a description of the stages of ECB as derived from case studies and narrative reports on various types of organizations. This is followed by a discussion of organizational change and the stages of change of organizations involved in ECB. A brief presentation of evaluation in the Canadian federal government follows in order to provide some context for the study. Some alternatives and critiques of ECB are then presented to balance the existing literature on ECB. Finally, some of the challenges of implementing and studying ECB initiatives are outlined to conclude the review. The literature review also provides an assessment of the quality of the knowledge base in the areas identified above by highlighting specific examples of published works on ECB.

Rationale for Evaluation Capacity Building

Service organizations such as government and non-profit institutions are increasingly making use of collaborative evaluation approaches managed by internal stakeholders rather than relying on periodic studies of individual programs conducted by external contractors (Sanders, 2002; Torres & Preskill, 2001). Underlying this fundamental shift in evaluation practice has been the recognition of the importance of participatory processes in fostering shared understandings of an organization's goals and achievements and in constructing new knowledge that can bring about organizational improvement (Cousins et al., 2004; Morabito, 2002; Owen & Lambert, 1998; Preskill & Torres, 1999, 2000; Preskill, Zuckerman, & Matthews, 2003; Robinson & Cousins, 2004; Rogers & Hough, 1995; Sanders, 2002; Torres & Preskill, 2001).

Collaborative approaches to evaluation and the development of an organization's evaluation capacity have certain commonalities. For instance, the primary purpose of Practical Participatory Evaluation (P-PE), like that of ECB, is to foster evaluation use through the direct participation of program stakeholders in evaluation activities. By involving stakeholders in a wide range of evaluation processes, P-PE aims to influence organizational decision-making based on evaluative evidence (Cousins & Whitmore, 1998). In their comprehensive review of the literature on ECB, Cousins et al. (2004) identify two types of ECB: direct ECB, which involves planned evaluation capacity building activities within an organization that occur either within or outside of actual evaluation projects (e.g., training on statistical data analysis) and indirect ECB, which results from the involvement of stakeholders in the evaluative process. In this latter sense, ECB is similar to participatory evaluation. However, ECB processes are different from participatory evaluation approaches in two ways: they are typically integrated into the organization's practices and they are ongoing rather than episodic or event-driven (Preskill &

Torres, 1999; Rowe & Jacobs, 1998; Stockdill et al., 2002). ECB processes have been linked to two consequences for organizations, evaluation use and organizational learning (Cousins et al., 2004).

Evaluation Use

Evaluation utilization in organizational decision-making is the main goal of many participatory evaluation approaches (Caracelli, 2000; Cousins, 1996; Cousins & Earl, 1992; Johnson, 1998; Morabito, 2002; Patton, 1999; Preskill & Torres, 1994; Robinson & Cousins, 2004; Torres & Preskill, 2001; Turnbull, 1999). The participation of program stakeholders in evaluation activities is critical in fostering shared understandings of a program's goals and achievements, and in creating new knowledge on organizational processes. Four types of evaluation use have been identified in the literature, the first three being associated with the use of findings: *instrumental use*, which refers to using evaluation findings as a basis for action and change; *conceptual use*, which occurs when an evaluation influences the stakeholders' understanding and attitudes about a program; *symbolic use*, which occurs when individuals use evaluation information for political self-interest. This type of use occurs when individuals have ulterior motives for their involvement with evaluation. For example, evaluation findings can be used to justify decisions already made about a program, or to advocate issues and persuade people to act (Johnson, 1998). Finally, *process use* is observed when behavioural and cognitive changes occur in evaluation stakeholder participants by virtue of their proximity to it. It occurs when individuals participate in evaluation activities and reflect upon these experiences. Process use moves beyond the use of findings and involves learning to apply evaluative logic and thinking to a broader program and organizational context (Alkin & Taut, 2002; Johnson, 1998; Patton, 1997). Although process use overlaps slightly with instrumental and conceptual use

(Alkin & Taut, 2002; Johnson, 1998), its focus on the development of new knowledge and skills related to evaluation practice sets it apart from the other types of use and has interesting implications in an ECB and organizational improvement context. Through process use, organizational 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).

In an empirical study focusing on the impact of process use on evaluation, Preskill, Zuckerman and Matthews (2003) identified five categories of variables that appear to influence process use in organizations: (1) facilitation of the evaluation process based on the creation of a learning environment, (2) management support for evaluation, (3) stakeholders’ interest in and knowledge of evaluation and the program under study, (4) frequency, methods, and quality of the communications occurring as part of the evaluation process, and (5) organizational stability and support for evaluation capacity building.

Newer forms of evaluation use have also been suggested in the literature. For instance, Weiss, Murphy-Graham and Birkeland (2005) have identified *imposed use* as an emerging form of evaluation utilization stemming from the context of increased accountability for government-funded initiatives. These researchers described imposed use as a situation in which “program stakeholders are obliged to pay attention to evaluation results” (p. 16). The example of imposed use provided in their paper refers to the influence of evaluation on the decisions made by various stakeholders to maintain or drop a drug prevention program offered in schools across the United States. In addition to these conceptualizations of use, the notion of evaluation influence has also garnered some interest in the recent literature. This term, first coined by Kirkhart (2000), attempts to move beyond the direct use of evaluation results and towards a more general

definition of evaluation use. Evaluation influence, according to Kirkhart's model, can be described according to three dimensions: source of influence, intention and time. Each of these dimensions can be further broken down into sub-dimensions. The notion of evaluation influence has been carried further by Henry and Mark (2003) and Mark and Henry (2004) in a framework developed to better understand this concept. Their framework describes evaluation influence on three levels: individual, interpersonal, and collective. At each of these levels, four to six mechanisms (or change processes) can occur as a result of an evaluation. Although the more traditional definition of evaluation use will be considered in this study, the direction taken by recent studies focusing exclusively on evaluation influence may prove to be an interesting path for future evaluation capacity studies (for example, see Christie, 2007).

Several factors have been shown to contribute to increased evaluation use. Some of these factors are related to the evaluation process itself, such as the role of the evaluator, the variety of evaluation methods used in the study, the adjustments made in the evaluation plan over the course of the study, and the involvement of program practitioners (Alkin & Taut, 2002; Conner, 1998; Cousins, 1996; Toulemonde, 1999). For example, in an empirical study of participation in evaluation and its relationship to evaluation use based on Structural Equation Modeling, Turnbull (1999) found that:

... participatory evaluation is likely to result in increased use if participants perceive that (a) their work place goals are participative; (b) they are able to participate to a desired degree; (c) they perceive that they have influence in the decision-making process; (d) they believe that the participatory process was efficacious in that it achieved its intended outcomes (p. 140).

Other factors thought to influence evaluation use are related to the context and structure of the organization, such as the number and function of program stakeholders involved in the study, the credibility of the participants in the process, and availability of administrative support (Alkin & Taut, 2002; Conner, 1998; Shulha & Cousins, 1997; Toulemonde, 1999). The organizational environment was shown to influence the use of evaluation results in a case study conducted by Russ-Eft, Atwood and Eggherman (2002) in a private sector organization. The authors found in this case that the organizational environment has considerable potential to influence instrumental use of evaluation findings. Interestingly, they also found that even when the organizational environment does not foster instrumental use, evidence of process use can still be found. In particular, they found that the evaluation process led to some process uses such as shared understandings, supporting and reinforcing the program intervention, increasing engagement, self-determination, and ownership.

Evaluation utilization is an important consequence of ECB initiatives because it is through the utilization of evaluation findings and processes that the likelihood of organizational improvement would increase. By learning about and participating in evaluation activities, program practitioners gain control over what evaluation questions are being asked, can better define the context of the program under study (i.e., can include information related to mitigating factors that may have an impact on the program's ability to attain its expected outcomes), and can offer relevant interpretations of the data collected over the course of the evaluation. They can also make better use of the evaluation findings and apply these to organizational problems (Forss, Samset & Cracknell, 1994; Preskill & Torres, 2000).

If stakeholders do indeed learn about evaluation, about the program being evaluated, and about each other from their engagement in the evaluation process, it can be said that

individual learning has taken place. However, if they share their learning with others in the organization, it is conceivable that team and/or organizational learning may occur as well. (Preskill et al., 2003, p. 424)

As with other areas of evaluation research, however, little empirical work exists on such topics as process use. Most of the literature published in this area is theoretical or anecdotal in nature, and offers little evidence or testing of these theories. Further, Leviton (2003) points out that even the empirical literature published on evaluation use relies heavily on self-reports of use of evaluative information, with no validation or triangulation of these reports. The knowledge base on evaluation utilization in general, and process use in particular, would benefit from further empirical studies that support currently accepted notions in these areas.

ECB initiatives therefore are thought to contribute to evaluation use, as well as to the development of a culture of systematic self-assessment and reflection in the organization (Cousins et al., 2004). This in turn may lead to increased organizational learning, which is the other main consequence of organizational evaluation capacity development.

Organizational Learning

Organizational learning (OL) can be defined as a process of growth and improvement through which past experiences are reflected upon and used as the basis for change (DiBella & Nevis, 1998; Huber, 1991; Kruse, 2003; Torres & Preskill, 2001; Vakola, 2000; Yeung, Ulrich, Nason, & Von Glinow, 1999). In other words, “organizational learning is the vehicle for utilizing past experiences, adapting to environmental changes and enabling future options” (Berends, Boersma, & Weggerman, 2003, p. 1036).

Organizational learning is the subject of a vast body of literature stemming from a number of academic disciplines in the social sciences. The way in which OL is approached by

various authors, therefore, depends on their own epistemological and methodological backgrounds. For example, Taylor (1947) viewed OL within the context of his scientific management theories and Argyris and Schon (1978) articulated their approach according to the principles of social psychology and the learning processes of individuals. In addition to this, there appears to be a division between practice-oriented literature, which focuses on the learning organization, and the academic literature, which focuses on organizational learning. Even though OL is defined in many different ways in the literature because of this diversity in discipline and focus, two major theories of OL can be found in the research literature. The first, called the normative perspective, suggests that learning, as a mechanism for organizational improvement, does not occur through chance or random action, but through the development and use of specific skills. Normative definitions of OL identify prerequisites which organizations must fulfill in order to be called “learning organizations” (Matlay, 2000). These theories focus on identifying the practices, systems and structures necessary in order for organizations to learn. Studies conducted in the realm of the normative perspective focus on organizational characteristics and processes likely to foster organizational learning.

The second theory of OL, the developmental perspective, is more descriptive in nature. It acknowledges that all organizations learn, consciously or otherwise, and that learning is linked to the organization’s lifecycle, without specific learning activities or interventions (DiBella & Nevis, 1998). This theoretical perspective posits that higher levels of learning are thought to be achieved as the organization matures and that the organization may take one of an unlimited number of paths in attaining higher levels of learning.

A third theory of organizational learning that is gaining popularity is the capability perspective suggested by DiBella and Nevis (1998). This theory recognizes that OL is ongoing

and focuses on the learning processes that already exist within the organization. In this perspective, all organizations have learning capabilities that are embedded in their structure and culture. There is no “best” approach for OL, rather, the focal point of organizational leaders in this area is to understand these learning processes and to use them to foster sustainable organizational change. This third approach differs from the normative perspective in that it uses each organization’s own characteristics to develop a customized organizational learning development plan. It also differs from the developmental perspective in that it takes a proactive, intentional stance towards organizational learning instead of waiting for it to occur naturally. Evaluation capacity building could be considered an organizational learning process within this third perspective, since it seeks organizational improvement through the implementation of specific, targeted initiatives adapted to an organization’s context and lifecycle.

The development of an organization’s evaluation capacity also represents one of the ways through which individual-level learning may be transferred to the organizational level (Berends, Boersma, & Weggeman, 2003; Popper & Lipshitz, 2000). Although an organization must first learn through its individual members, OL is fundamentally different from individual learning and occurs through the social construction of knowledge to which those involved contribute; the evaluation process, therefore, can frame the interactions of organizational members and be the basis for knowledge construction (Klimecki & Lasseben, 1998; Jenlink, 1994; Preskill & Torres, 1999, 2000; Wenger, 2004).

Along the same lines, an increase in evaluation capacity may help organizations move beyond single-loop (or instrumental) learning and into double-loop learning (Argyris & Schon, 1978). Double-loop learning focuses on revisiting tightly held assumptions and beliefs about organizational goals and is closer to the definition of conceptual evaluation use (Alkin & Taut,

2002; Forss et al., 1994; Preskill & Torres, 1999). It has also been suggested that institutionalized evaluation practices, sometimes referred to as organizational evaluative inquiry, can contribute to deuterio learning, or learning how to learn, by providing organizational members with the tools and skills required to learn from past organizational problems (Torres & Preskill, 1999). Once again, however, little empirical evidence exists to support these claims since most of the literature in this area is theoretical in nature. Although useful at a conceptual level, further empirical work will lead to a greater understanding of OL and its manifestations in organizations.

In order for evaluation use and organizational learning to occur as a result of evaluation capacity building, several factors or conditions have been identified in the literature and are helpful in drawing a picture of how evaluation capacity might be developed in an organization. The following section describes these factors.

Organizational Factors Contributing to the Success of ECB

A number of factors or conditions leading to successful evaluation capacity building in organizations have been identified in recent years. In order to clarify and organize these factors, I have classified them into the four categories outlined below. These four categories were chosen based on existing categories found in the literature, as well as on my own analysis of the individual factors identified in various research articles. They are inclusive of all factors identified thus far in the literature.

External Environment

The external environment of the organization influences the successful implementation of ECB initiatives through the demands for evaluation that it places on the organization, the

incentives that it provides for conducting evaluation, and through the interactions that occur between external stakeholders and organizational members on matters of evaluation.

Demand and incentives. In order for ECB to be implemented in a sustainable manner, the external environment of the organization must create a demand for evaluation results (Stockdill et al., 2002). External accountability requirements constitute a powerful motivator for developing evaluation capacity by highlighting the importance of evidence-based decision making (Katz, Sutherland, & Earl, 2002; Mackay, 2002; Preskill et al., 2003; Sutherland, 2004; Toulemonde, 1999). These requirements can also influence evaluation behaviour within an organization by setting a baseline for the amount and type of evaluation activity to be conducted (Gibbs, Napp, Jolly, Westover & Uhl, 2002). Aside from creating a demand for evaluation results, the external environment of the organization can also offer incentives to foster evaluation capacity building such as increased financial resources for the organization and performance bonuses for managers (Boyle et al., 1999; Toulemonde, 1999; Trevisan, 2002).

Interaction with other organizations. The interaction that occurs between an organization and its environment may also result in increased ECB within the organization, even if no specific demand or incentives are present. The evaluation of large-scale, jointly managed programs that occurs in the government sector is one such example. Various organizations must cooperate in the evaluation of these programs, and these organizations may choose to increase their evaluation capacity in order to properly represent their own interests over the course of the evaluation (Milstein, Chapel, Wetterhall, & Cotton, 2002; Rowe & Jacobs, 1998).

The demand for evaluation services as well as the organization's interaction with other organizations may act as forces that determine the extent to which the organization will be motivated to develop its evaluation capacity. When studied from the perspective of institutional

theory, these environmental forces act to push the organization towards the achievement of greater legitimacy (Scott, 2001). Further details about how institutional theory can help us to analyze issues related to evaluation capacity building are provided in Chapter 7 of this thesis.

Organizational Structure

This category of factors refers to the organizational design, the available resources for evaluation in the organization, and the evaluation processes already in use.

Organizational design. The systems and staffing structures of organizations mediate organizational members' ability to interact, collaborate, and communicate with each other (Preskill & Torres, 2000). Organizational roles must be flexible enough to allow people to step away from their main responsibilities and participate in evaluation activities whenever required (Torres & Preskill, 2001). The assignment of organizational members to evaluation-related projects must take into consideration certain individual characteristics such as knowledge of the program being evaluated, interest in evaluation, and job level, in order to ensure that the right people are involved in evaluative inquiry (Preskill & Torres, 1999; Preskill et al., 2003; Sonnichsen, 1999).

The location of the evaluation function within the organization is also a key factor in organizational design. The presence of an internal evaluation function appears to be the best mechanism through which to ensure the integration of evaluation activities within program management (Caracelli, 2000; Cousins & Earl, 1992; King, 2002; Sonnichsen, 1999). Compared to external evaluators brought into the organization to work on specific projects for a limited period of time, internal evaluators are better able to facilitate environments that are conducive to learning, to act as change agents focused on organizational improvement and development, and

to empower their organization (Compton, Glover-Kudon, Smith, & Avery, 2002; Morabito, 2002; Sonnichsen, 1999; Torres & Preskill, 2001; Trevisan, 2002).

The internal evaluation function can be either centralized in a corporate office, or decentralized and distributed among functional areas. A centralized function is preferable when the main purpose of evaluation is accountability because it increases the independence of evaluators and allows for better coordination of evaluation activities across the organization (Compton et al., 2002; Milstein et al., 2002; Sonnichsen, 1999). Decentralized evaluation refers to the inclusion of an evaluation position within each function or program area of an organization. The evaluator is a member of the program team and links the evaluation function to the program function. This type of structure is more conducive to formative evaluation and is more likely to lead to long-term, sustainable evaluation capacity when a concrete ECB plan is in place (Caracelli, 2000; Jenlink, 1994; King, 2002; Milstein et al., 2002; Preskill, 1994; Sonnichsen, 1999).

Resources available for evaluative inquiry. Another factor related to organizational structure is that of the human and financial resources allocated to evaluative inquiry in the organization. Evaluation-related human resources are required to develop and implement the ECB strategy, to guide program practitioners through evaluation activities, and to provide specialized technical skills when required. Other human resources are also required to relieve program practitioners from their regular duties while they participate in evaluation activities, as well as to implement evaluation findings (Bozzo, 2002; Cousins & Earl, 1992; King, 2002; Milstein et al., 2002; Stevenson, Floris, Mills, & Andrade, 2002; Stockdill et al., 2002; Trevisan, 2002). Financial resources are required to support the training of practitioners in evaluation, to purchase specialized equipment or software needed for evaluation studies, and to disseminate

evaluation findings to key stakeholders (Boyle et al., 1999; Gibbs et al., 2002; King, 2002; Rowe & Jacobs, 1998; Sonnichsen, 1999; Trevisan, 2002; Toulemonde, 1999).

Evaluation processes within the organization. A third factor related to organizational structure identified in the literature focuses on the evaluation processes that exist within the organization. As stated previously, many of the characteristics associated with successful participatory evaluation are a good starting point for ECB; however, building an organization's evaluation capacity requires more than a participatory approach if the goal is to make evaluation practice common to all levels of the organization. For instance, like in participatory approaches, the evaluation practices used in ECB must be conducive to learning by identifying evaluation issues that are relevant and easy to understand by organizational members and by using a methodology that is acceptable and credible to program staff and that fits within the organizational culture and structure (Compton et al., 2002; Cousins & Earl, 1992; Sonnichsen, 1999). Beyond this, the evaluation activities must also include a variety of organizational members and must focus on various initiatives, programs, and policies in an integrated manner (Torres & Preskill, 2001).

The adoption of a conceptual framework of evaluation may also help increase the likelihood of continued evaluation practice across the entire organization. One example of this is provided by Compton et al. (2002) in their narrative of ECB implementation efforts at the American Cancer Society (ACS). According to these authors, such a framework provided the ACS with a useful reference summarizing the purposes of evaluation, the evaluation process, the uses of evaluation, and an evaluation vocabulary for non-evaluators.

Organizational Culture

The third category of factors conducive to ECB refers to the existence of a supportive organizational setting that influences individual and organizational learning and performance (Preskill, 1994; Preskill & Torres, 1999; Trevisan, 2002). The culture of an organization reflects the traditions, values, and basic assumptions shared by its members and that establish its behavioural norms. This includes developing clear goals and objectives as well as fostering a learning climate.

Clear goals and objectives. In order to be successful, ECB initiatives must be intentional and strategic. They must be implemented in stages so that program practitioners are not overwhelmed with information and activities; consequently, ECB initiatives require clear goals and objectives (Duignan, 2003; King, 2002). It is critical that organizational members understand the ECB plan and how it supports the goals of the organization. This will help evaluators avoid unnecessary surprises and respond to the organization's learning needs efficiently (Sonnichsen, 1999; Stockdill et al., 2002). An interesting example of the importance of implementing ECB by following a planned approach is provided by King (2002). This case narrative describes the efforts undertaken by an evaluator to build the capacity of a school district to do and use evaluation. The ECB plan implemented focused on four goals: (1) develop staff commitment and skills in program evaluation and its use; (2) build an infrastructure for data collection, analysis and presentation to support evaluation and its use; (3) facilitate the existing school improvement process, and (4) create a network of people across the district to routinize organizational inquiry. Even though the organizational context and budget cuts made it difficult in this case to implement ECB according to the plan articulated at the outset, the stated goals of the ECB efforts

remain a key commitment of organizational leaders and represent a common vision for the school district's future.

The organization itself must also have a clear mission and clear objectives that can be measured through evaluation activities (Rowe & Jacobs, 1998; Goh, 2003). This is crucial to the sustainable integration of evaluative inquiry across all levels of the organization.

Fostering a learning climate. Organizations must have the proper climate in order to be able to participate fully in evaluation processes and to use the findings for organizational improvement (Boyle et al., 1999; Russ-Eft et al., 2002). This involves creating an organizational setting in which it is safe and rewarded to question organizational processes and experiment with new approaches (Goh, 2003; Preskill et al., 2003; Preskill & Torres, 1999; Rowe & Jacobs, 1998; Torres & Preskill, 2001; Toulemonde, 1999). Although organizational leaders play a critical role in fostering supportive climates for ECB, the organization's culture may, in turn, help convince leaders to adopt evaluation practices when they otherwise would not support such efforts by providing "the collective pressure that makes decision makers overcome their reluctance, even when evaluation deeply contradicts their self-interest" (Toulemonde, 1999, p. 167). The role of organizational leaders in supporting ECB initiatives is described in the following section.

Organizational Leadership

Managerial support is necessary to the implementation and sustainability of evaluation capacity within an organization. In order for ECB initiatives to succeed, organizational leaders must be committed to evaluation and believe in its potential to improve the organization's effectiveness while balancing the demand for accountability from the organization's external environment (Goh & Richards, 1997; Goh, 2003; King, 2002; Milstein et al., 2002). Senior

managers must make evaluative inquiry a priority by allowing staff the flexibility required to participate in ECB activities while fostering a climate of risk-taking and experimentation within the organization. They must also be advocates for learning in the organization, and they must support and reward participation in evaluation work and in the implementation of evaluation results (Gibbs et al., 2002; Goh, 2003; Leonard & Miller, 1992; Preskill et al., 2003). It is also critical that organizational leaders understand the time frame required to establish evaluation capacity and that they develop and implement a clear strategy to accomplish this (Boyle et al., 1999; Milstein et al., 2002). Gibbs et al. (2002) underline the importance of strong leadership in their analysis of community-based ECB initiatives. These authors found that organizations whose leaders believed that evaluation could be used to improve program effectiveness and that strategies could be found to overcome the many challenges of evaluation conducted more extensive or complex evaluations than had been required of them by external funding agencies. Milstein et al. (2002) support this further in their case narrative summarizing their efforts towards building evaluation capacity at the Centers for Disease Control (CDC). These authors outline the role of senior managers in establishing program evaluation as a priority for the organization and the effects of ECB on the organization. The process used to build evaluation capacity throughout the CDC are related in this paper, which concludes, among other things, that: “To build capacity, agency leaders must convince staff and stakeholders that answering evaluation questions thoroughly and honestly is more important than getting results that look and feel good” (p. 42).

In addition to conceptualizing evaluation capacity building as the outcome of a discrete series of factors, it is also possible to view it within the broader spectrum of organizational

improvement. The following section positions evaluation capacity building within this broader context in order to further our understanding of its key characteristics.

Planned Organizational Change

ECB can be considered a mechanism through which organizational innovation and change occur. The organizational change process can generally be defined as a cycle of knowledge creation, dissemination, and adoption (Johnson, 1998). Institutional theorists, in particular, define change as a difference in the form, quality, or state over time in an organization's relationship or alignment with its environment (Fletcher, 1990; Rajagopalan & Spreitzer, 1997; Van de Ven & Poole, 1995). According to institutional theory, the change process typically begins with the identification of a discrepancy between the organization's internal and external environments and ends once a new process has been put in place across the organization to deal with this discrepancy and has been fully adopted by organizational members (Van de Ven, 1993).

Such an emphasis on the external environment as the main influence on an organization's propensity and ability to change, however, has received some criticism. Opponents of this view argue that the environment's role has been overemphasized in terms of directing the change process and state that although the environment may have some influence over an organizational decision to implement change, other factors must also be considered at the organizational level and may have more weight than environmental pressures. Critics also note that by selecting the organization as unit of analysis, researchers have given little credence to individual agency and strategic choice, factors considered important in organizational decision-making. In other words, studies conducted on organizational change thus far have examined large-scale changes, but do

not present the level of detail that may be necessary for a comprehensive understanding of organizational change (Stevens & Slack, 1998).

In ECB terms, organizational change amounts to a perceived need for timely and useful evaluation information that meets the needs of program managers and other decision makers, usually prompted by external evaluation requirements. Evaluation, as a starting point for organizational change, is often reactive rather than proactive by its very nature – programs are developed, implemented, then evaluated. One exception to this is needs assessment, which is considered part of the evaluator's tool kit. Needs assessments are conducted prior to program design and provide information on what program elements are required.

Barriers and Facilitators to Organizational Change

Several factors have been found to facilitate or inhibit organizational change initiatives. These include factors related to an organization's structure, such as its size, the flexibility of the organizational roles, and its capacity to manage change while conducting its regular business (Fullan, 1999; Johnson, 1998; Rajagopalan & Spreitzer, 1997). More importantly, however, the success of organizational change depends on an organization's ability to learn. Successful learning organizations can anticipate and react to changing circumstances in a proactive manner and have the required self-knowledge to identify a need for change as it arises (Hill, Hazlett, & Meegan, 2001; Johnson, 1998). The culture of learning organizations, which supports risk-taking and experimenting, also acts as an enabler of organizational change by speeding up institutionalization and by questioning the organization's deeply held values and assumptions (Comfort, 1997; Fullan, 1999; Inhetveen, 1999; Payne, 1996).

Models of Organizational Change

Organizational change has been the focus of a multitude of theories and models, all attempting to describe the change process and predict the success of future change initiatives. These theories and models can be slotted into one of six general categories: evolutionary, teleological, life cycle, dialectical, social cognition, and cultural (Kezar, 2001; Van de Ven & Poole, 1995). Although these broad categories are helpful in bringing forward our conceptual understanding of organizational change, these models have been criticized in terms of their simplicity and lack of applicability to empirical situations (Frankreiss & Victor, 2002; Prochaska, Prochaska, & Levesque, 2001).

Models with a different orientation than the six general categories discussed above have also been developed by organizational theorists and offer a new perspective on the organizational change process. Different types of change, for instance, have been proposed, such as incremental change versus discontinuous change, or reactive change versus anticipatory change (Poole, 1996). Other authors have developed general principles of change, rather than focusing on conceptual relationships (Frankreiss & Victor, 2002; Miller, Lant, Milliken, & Korn., 1996; Rajagopalan & Spreitzer, 1997). For these researchers, change does not always follow the same pattern from one organization to the other, but the change experience has several elements that are applicable across certain organizational contexts. Rajagopalan and Spreitzer (1997) speak of change as defined by three lenses, rational, learning, or cognitive, depending on the organization. The rational lens defines change as a sequential, planned search for optimal solutions for clearly identified problems, based on previously defined organizational objectives. The learning lens views change as an iterative process, as organizational participants learn from their experiences. Finally, the cognitive lens emphasizes the interpretive processes through which managers

interact with the environmental and organizational contexts. Miller et al. (1996) focus on organizational adaptation rather than change, and identify two models within this area: the passive model of adaptation, characterized by routine-driven, incremental behaviour that focuses on organizational stability, and the opportunistic model of adaptation, characterized by the search for, and exploration of, a wide variety of alternative goals, activities, and modes of operation.

Educational organizations such as schools have been the subject of several organizational change models. Hannay and Ross (1997), for instance, have developed three categories that describe schools engaged in organizational change: waiting schools, hovering schools, and initiating schools. According to this model, waiting schools have made some superficial changes but appear to be waiting to decide their direction based on a variety of factors; hovering schools have made more changes regarding their structure, but are not entirely committed to the change initiative; finally, initiating schools demonstrate fundamental change and question the former structure. Leithwood and Montgomery (1987) present a more specific framework focusing on planned change of classroom practice. The key concepts of this framework include incremental growth, individual roles within the broader system of the school, and obstacles and strategies to growth.

The literature on innovation and the adaptation of organizations to innovative practices is also replete with models of organizational change. Rogers' (1995) model of diffusion of innovations is perhaps the most popular view of the organizational change process within a context of innovation adoption and offers an interesting approach through which ECB can be considered. This model portrays the process of innovation over time as a linear sequence of three stages, beginning with the invention of an idea, through its development, production and testing, and ending with its diffusion to and adaptation by users. Other researchers have developed

Rogers' original idea further and identified several levels of use of an innovation and the steps that lead to the adoption of an innovation by an individual or by an organization (Hall, Loucks, Rutherford, & Newlove, 1975).

Finally, some researchers have attempted to integrate a number of theoretical constructs in order to develop a general theory of organizational change applicable across contexts. One such example is the Transtheoretical Model (TTM), which integrates constructs such as stages of change, decisional balance, and processes of change. This model promotes the use of stage-matched interventions at an individual level to reduce resistance, stress, and the time to implement the planned change (Prochaska, Prochaska, & Levesque, 2001).

In summary, the organizational change literature identifies several different types of organizational change models, each focusing on a different aspect of the change process, such as the influence of the environment, the adaptation to innovation, or the type of organization in which change takes place. Most of these models follow a staged approach to change; some of these approaches are described in the following section.

Stages of Change

Staged approaches to change are typically organized in three major steps: the organization's current state, the desired future state, and the transitional phase in which change initiatives are implemented and institutionalized (Hill, Hazlett & Meegan, 2001). The transitional phase can be further broken down into discrete steps depending on the type of change to be implemented. Some change models are articulated around a time-based approach, such as the Transtheoretical Model put forward by Prochaska et al. (2001). This model identifies five stages through which people or organizations progress; each stage occurs at a specific time, starting up to one year before change initiatives begin, and ending 6 months after the change is

implemented. Most staged approaches, however, focus on a number of qualitatively different steps organized in a linear model, such as those developed by Carnall (2003), Hall et al. (1975), Loucks-Horsley & Hergert (1985), and Rogers (1995). Rogers' model is perhaps most interesting since unlike the other models, it uses terminology specific to organizations rather than extrapolating individual-level stages of change to the broader context of organizational change. For instance, Rogers (1995) speaks of "agenda-setting" as the first stage of change, while Buckley and Perkins (1985; cited in Fletcher, 1990), identify the first stage as "unconsciousness", and Carnall (1993) calls it "denial".

Because of its organizational orientation and terminology, Rogers' model suggests a useful approach through which to consider ECB in organizations. The first stage of this approach, "agenda-setting", occurs when a general organizational problem that may create a need for an innovation is defined. This stage includes a search of the organization's environment to locate innovations of potential usefulness to meet the organization's problem. The second stage, "matching", occurs when a problem from the organization's agenda is fit with an innovation, and this match is planned and designed. At this stage, the problem is conceptually matched with the innovation to establish how well they fit together. In the third stage, "redefining and restructuring", the innovation is imported from the external environment and is re-invented in order to accommodate the organization's needs and structure most closely. At the same time, the structure of the organization is also modified to fit with the innovation. As the innovation is put into widespread use in the organization, the fourth stage, "clarifying", occurs. In this stage, the innovation becomes imbedded in the organization's structure, and misunderstandings or unwanted side-effects of the innovation occur and are resolved through communication and interaction between organizational members. Finally, "routinizing" occurs

when the innovation has become incorporated into the regular activities of the organization and is no longer considered an innovation separate from the other organizational processes and procedures. In other words, routinizing occurs when organizational members no longer think of the innovation as a new idea. Another model, the seven-step school improvement process outlined by Loucks-Horsley and Hergert (1985), covers the same basic steps as the model developed by Rogers, but adds two additional steps, establishing the school improvement process at the outset of the change initiative and reviewing progress and problems once routinizing has been achieved. The first additional step involves negotiating for resources, building a base of relationships and forming a school improvement team, and the last stage includes analyzing progress and perceptions through monitoring, evaluating outcomes, and making further refinements.

Stages of Evaluation Capacity Building

Transitioning to an organizational setting in which evaluation is a predominant factor in organizational structure and decision-making requires intentional, continuous effort (Baizerman, Compton, & Stockdill, 2002). Although very few authors currently conceptualize ECB in terms of stages of development, the use of a developmental approach is helpful in understanding ECB as a growth process occurring in organizations. Therefore, as a starting point, I have identified four tentative stages of ECB based on some of the general stages outlined in the literature and on an analysis of more qualitative descriptions of the evaluation capacity building efforts of specific organizations.

The first stage, *traditional evaluation*, is the initial state of ECB. At this stage, evaluation is entirely mandated by external factors; individual programs are evaluated as required, often by external contractors hired specifically for this purpose. This stage is also characterized by

minimal involvement on the part of program stakeholders (Gilliam, Barrington, Davis, Lacson, Uhl & Phoenix, 2003; Sanders, 2002; Torres & Preskill, 2001). It is similar to the stage of *compliance* identified by Gibbs et al. (2002) in their case study of ECB implementation in community-based organizations dealing with HIV prevention programs. These authors characterize this stage as evaluation conducted only to the extent required by external bodies with the perception that evaluation brings no benefit to the organization other than continued funding.

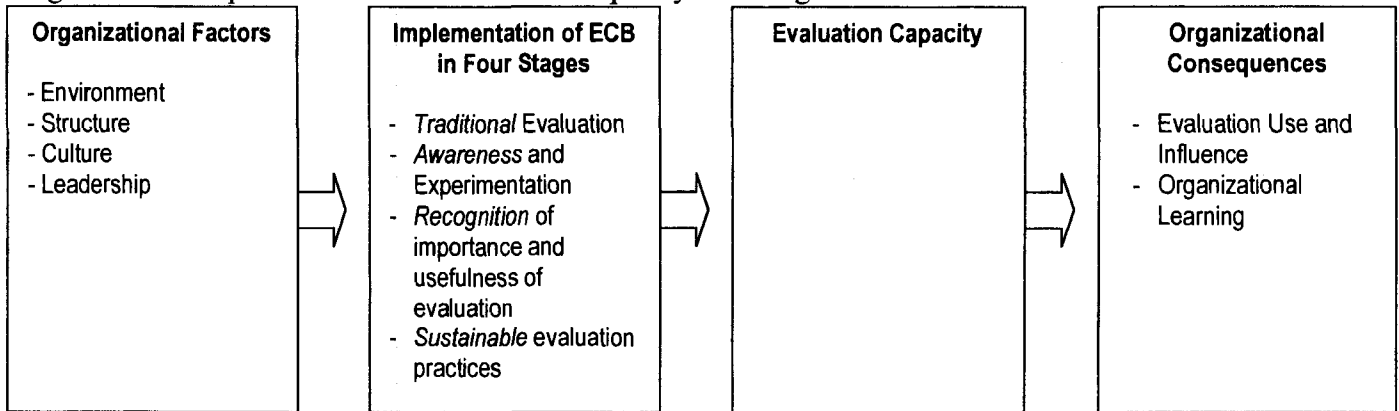
The second stage, *awareness and experimentation*, begins when organizational members become aware of a need to change certain activities, usually because of changes occurring within the external environment of the organization. This stage involves collecting and analyzing data through more participatory evaluation approaches and leads to a greater understanding of program goals and outcomes (Sanders, 2002; Torres & Preskill, 2001). Once organizations experience success with these participatory evaluation approaches, they may begin to *implement evaluative inquiry* throughout the organization. Specific plans are developed and resources are committed to ECB initiatives at this stage and the organizational culture starts to move towards greater acceptance of systematic reflection on practices and procedures (Duignan, 2003; Gibbs et al., 2002; Sanders, 2002; Torres & Preskill, 2001).

Finally, the organization *adopts evaluative inquiry as a management function and sustains its practice*, regardless of other internal or external constraints. Evaluative inquiry becomes part of the tightly held values of the organization and new members are initiated into this culture of self-reflection and learning (Gibbs et al., 2002; Lewis & Thornhill, 1994; Sanders, 2002; Torres & Preskill, 2001).

Summary: A Conceptual Model of Evaluation Capacity Building

A graphic description of the main concepts associated with evaluation capacity building may be helpful in summarizing the literature in this area. Figure 1 has been developed for this purpose and is presented below.

Figure 1: Conceptual Model of Evaluation Capacity Building



As can be seen in Figure 1, much is known about the organizational factors necessary to the development of evaluation capacity. The external environment of the organization as well as its structure, culture and leadership, have been shown both anecdotally and empirically to influence the success of ECB efforts. Once evaluation capacity building initiatives are undertaken, organizations seem to follow a similar path (as outlined in anecdotal and empirical reports of ECB initiatives), from traditional evaluation practices to sustainable evaluation inquiry at all levels of the organization. The different stages of ECB are also well documented in the evaluation literature and parallel those identified in the organizational change literature. Finally, the literature also identifies and describes the two main consequences of improved evaluation capacity, evaluation use and organizational learning. However, the intermediate step in the diagram, the attainment of evaluation capacity and what that might look like at a practical level, is not well described in the literature and is therefore the focus of the present study. As

mentioned earlier, evaluation capacity and ECB are closely intertwined and have only been separated here for heuristic purposes. Evaluation capacity is conceptualized here as a component of the broader conceptual framework describing evaluation capacity building. Since much is already known about the other components of ECB, evaluation capacity has been split apart from the more general concept in order to be studied more closely in this thesis.

Evaluation in the Canadian Federal Government

Program evaluation was first officially implemented in the Canadian federal government in 1977, when the first policy on evaluation was published (Segsworth, 2005). It was expected at that time that independent and objective evaluation studies would inform senior decision makers by focusing on issues of relevance, success, and cost-effectiveness. All programs were to be evaluated on a three- to five-year cycle, and the methods used were taken from the fields of education and health because of their strong academic and evaluation backgrounds (Segsworth, 2005; Young, 2006).

Five audits of the evaluation function in the federal government have taken place since 1977. Overall, the findings of these audits have indicated that evaluation has not lived up to its potential and provides little useful information to senior officials (McDavid & Huse, 2006). Evaluation has evolved slowly and recent criticisms point to its focus on internal, formative issues rather than on broad-reaching, strategic issues (Hunt, 2006). Although there have been some examples of success along the way, problems with quality and credibility have tainted the evaluation function in the eyes of senior managers (McDavid & Huse, 2006; Young, 2006).

The introduction of results-based management in the early 2000s has given rise to a new opportunity for government evaluators. In some departments, this new focus has been embraced and a transition towards offering performance measurement advice and services and has resulted

in a renewed usefulness at all levels of the organization (McDavid & Huse, 2006; Young, 2006). It has been argued that in order to remain relevant, "...there will need to be a movement (for evaluation) away from a singular focus on producing more and more one-off evaluations and toward a synthesis and integration of evaluative knowledge into management practices and policies" (Mayne & Rist, 2006, p. 94-95). Although performance measurement should remain the responsibility of program managers, evaluators can provide expertise not found elsewhere in the organization and in this way, contribute to senior-level decision-making in their organizations (Maybe, 2006). Other authors, however, caution against the heavy involvement of evaluators in performance measurement and away from traditional evaluation studies: although performance measurement can assist evaluators by clarifying policy intent and program goals and by facilitating data collection, it cannot serve as a substitute for evaluation because it does not make claims about a program's merit or worth (Mohan, Tikoo, Capela & Bernstein, 2006).

Evaluators therefore appear to have two distinct roles in influencing policy decisions: one as advocates and supporters of performance measurement¹, and the other as providers of evaluation information. Because government is an inherently political arena, evaluators must be adept at managing the politics of evaluation in order to have an impact on senior-level decision-making. This involves understanding the evaluation environment and maintaining both independence and responsiveness throughout the evaluation process (Mohan & Sullivan, 2006). Encouraging evaluation capacity building within their organization while maintaining the independence required to produce credible studies is a challenging undertaking. For this reason and others, some alternatives to ECB have been suggested in the literature. These are considered

¹ Performance measurement, also known as performance management, refers to the ongoing collection of program data for program management purposes. It differs from evaluation in that it is continuous rather than episodic, and focuses on concrete indicators and measures related to program delivery and early outcomes.

in the following section in an attempt to better understand the criticisms that have been made of ECB in government and how these may be remedied in the future.

Alternatives to Evaluation Capacity Building

Two main alternatives to ECB have been suggested in the literature. Although neither has been suggested specifically within the context of government evaluation, both could be applied in this context.

Mainstreaming

Mainstreaming has been defined as “the process of making evaluation an integral part of an organization’s everyday operations” (Sanders, 2003, p. 3). Like evaluation capacity building, its focus is organizational improvement through the construction of shared knowledge.

Mainstreaming is thought to be a partial result of ECB initiatives and moves evaluation to the forefront of organizational thinking and behaviour. Its proponents distinguish it from ECB by defining ECB as an array of practices undertaken on an episodic basis to improve the quality and use of evaluation in an organization (Duignan, 2002; Sanders, 2002, 2003). However, increasingly accepted definitions of ECB (such as the one used in this study, from Stockdill et al., 2002) include elements of continuous rather than episodic efforts focusing on incorporating evaluative thinking throughout the organization. In this sense, mainstreaming and ECB represent similar strategies for organizational improvement.

The results of a literature search undertaken to identify papers on mainstreaming show that it was considered a topic of interest in the years 2002 to 2004, with eight publications appearing in evaluation journals between these years, and none subsequently. This may be indicative of the fact that mainstreaming, by and large, has been replaced by the more extensive definition of evaluation capacity building.

Insourcing

A recent critique of ECB suggests that insourcing may be a better model for community-based organizations with limited resources for staff and evaluative activities (Miller, Kobayashi, & Noble, 2006). Instead of training staff members to conduct evaluation as in ECB or to hire external evaluators, the insourcing model offers a mix of the two by featuring data collection conducted by staff and analysis done by external evaluators. This model focuses mainly on the collection of quantitative data undertaken in an effort to track specific performance indicators. It relies on the services of an external evaluator hired by a group of like-minded organizations; the evaluator works with each of the organizations to develop a performance measurement plan articulated around a logic model and customizes the necessary data collection tools by using existing templates. Staff members are engaged infrequently to collect data, to discuss evaluation results and to provide feedback on measurement issues and program direction.

The creators of the insourcing model provide a few examples to illustrate how the model works in practice, but no empirical evidence is offered to demonstrate that it is a better alternative to ECB (Miller et al., 2006). The model assumes that the evaluation results generated by the external evaluator will be relevant and helpful to program practitioners, even though their involvement in the evaluation design process is minimal. In addition to this, unlike ECB, no linkage, theoretical or empirical, has been made between organizational learning and insourcing. In this sense, insourcing does not constitute a viable alternative to ECB.

A similar approach to insourcing that is clearly within the realm of ECB is that used by Jolly, Gibbs, Napp, Westover, & Uhl (2003). These authors conducted an empirical qualitative study in order to identify the types of technical assistance required by community-based organizations, to understand these organizations' past experiences with evaluation, and to

generate ideas for the best possible delivery of evaluation technical assistance. The study findings point to the fact that community-based organizations (in this case, organizations focusing on HIV prevention programs) require technical assistance in familiarizing organizational members with basic evaluation terminology and processes, planning and designing evaluations, designing or identifying appropriate evaluation tools, data analysis, and becoming knowledgeable partners in evaluation. Although these are consistent to some extent with the gaps filled by insourcing, the type of technical assistance required here clearly moves beyond instrumental and conceptual uses of evaluation and focuses more specifically on developing process use, a key component of evaluation capacity building.

Theoretical Critique of ECB

The rapid growth of the pragmatic knowledge base on ECB in recent years has left little room in the literature for more theoretical considerations and systematic critiques of this approach and the underlying concept of evaluation capacity. The literature on ECB assumes that evaluation in general is a worthwhile pursuit, and that those organizations that have a well-developed evaluation capacity benefit from the evaluation activities undertaken by staff members. Because program evaluation has traditionally been conducted in service organizations, there is also an assumption in the literature that the goals espoused by these organizations are worthwhile and attainable. This section explores these assumptions as they pertain to public service institutions through the lens of Foucault's conception of governmentality and as they relate to workplace learning through his views on disciplinary power. Both of these theoretical constructs offer a counterbalance to the knowledge base on ECB and will be of use in the analysis and interpretation of the data collected in the proposed study.

Governmentality and Evaluation

The governmentality approach defines government as:

any more or less calculated and rational activity, undertaken by a multiplicity of authorities and agencies, employing a variety of techniques and forms of knowledge, that seeks to shape conduct by working through our desires, aspirations, interests and beliefs, for definite but shifting ends and with a diverse set of relatively unpredictable consequences, effects and outcomes (Dean, 1999, p. 11).

Key to this definition are the notions of conduct-shaping, outcomes and techniques.

Government has been referred to as “the conduct of conduct” (Colebatch, 2002). In other words, governmental institutions are thought to direct the actions of individual citizens through incentives, demands, and consequences delivered in the form of governmental interventions, or programs. The issue of power is central to government programming, since it is through their programs that governmental institutions exert their power, in an attempt to attain specific outcomes (Rose & Miller, 1992)

The role of evaluation within this framework is to ensure that the *techniques* employed by government to *shape the conduct* of individuals do indeed lead to the desired *outcomes*.

Evaluation and its counterparts (i.e., performance measurement, strategic planning, etc.) can be thought of as “technologies of performance” (Dean, 1999), or a “more or less technical means for locking the moral and political requirements of the shaping of the conduct into the optimization of performance” (Dean, 1999, p. 169). This definition describes well the major concern associated with evaluation practice, i.e., that it can sometimes miss the more important questions of relevance and appropriateness of government intervention in certain areas by focusing too closely on technical matters pertaining to the achievement of specific outcomes

through prescribed means in an audit-like fashion. In this sense, evaluation can be seen as a mechanism through which status quo is preserved and basic assumptions are never explored. While this may be true to some extent of traditional evaluation, the more participatory forms of evaluation espoused by organizations involved in ECB provide a setting in which it is appropriate for stakeholders to question the overarching goals of government intervention and the techniques used for conduct-shaping. However, even these participatory approaches to evaluation are susceptible to power imbalances such as those described by Foucault and his concept of disciplinary power.

Disciplinary Power and the Evaluator

The issue of participation central to evaluation capacity building can also be viewed through the lens of Foucault's conceptualization of power. Governmentality focuses on the power of institutions on individual citizens external to these institutions. The notion of disciplinary power is more applicable to interpersonal relationships and is particularly interesting in a workplace education context: students exercise as much power over each other and the educator as the educator exercises over them (Brookfield, 2001). The application of this perspective to ECB is helpful in identifying the different roles taken by evaluators charged with building the capacity of organizations to do and use evaluation.

Morabito (2000) identifies four roles for evaluators engaged in ECB: educator, consultant, counselor, and facilitator. Each of these roles requires the evaluator to reflect upon power and how it influences the process of evaluation capacity building in the organization. For example, the *evaluator as educator* must be aware of the power distribution within the group of participants and use it productively. The *evaluator as consultant* holds some power over program practitioners since he or she has evaluation-related skills and knowledge that the others do not

have. The evaluator therefore has full use of the “dominant discourse” of evaluation, and controls access to this discourse (Brookfield, 2001). The *evaluator as counselor* “seeks to empower the organization through exploring values, defining problems, gathering information, and fostering organizational change and learning based on its needs” (Morabito, 2002, p. 327). This implies that the organization speaks with one voice and that it is possible to ‘empower’ the entire organization through ECB. However, Foucault has taught us that empowering some members of an organization will likely result in reducing the power of others, and evaluators must be aware of this throughout the ECB process. Finally, the *evaluator as facilitator* must make the process as smooth and easy as possible in order for people to feel comfortable sharing their views and discussing them openly with others. Depending on the organizational setting and the evaluation group’s composition, this may be difficult to accomplish.

The theoretical grounding of ECB within Foucault’s conceptions of governmentality and disciplinary power provides a critique of ECB not found elsewhere in this relatively new body of literature. The frameworks within which Foucault operates have served as useful reminders throughout the identification of the dimensions of evaluation capacity that is the central focus of the thesis as well as throughout the development of the framework that operationalizes these dimensions by bringing to the forefront the normative nature of organizations and contributing to a deeper understanding of the data collected through the interviews with organizational members.

Profiling Organizational Growth in Evaluation Capacity

As the body of literature on ECB has grown, so has our understanding of the factors that influence it and the stages through which organizations may move as they develop their capacity to do and use evaluation. However, as demonstrated in the conceptual model presented

previously, little is currently known about the specific dimensions of evaluation capacity in organizations and how these may be identified and observed (as demonstrated by the absence of content in the 'evaluation capacity' section of the conceptual model). The Innovation Profiles method developed by Leithwood and Montgomery (1987) features a framework for planned change of classroom practices and will be used as a basis for the framework that has been developed as a result of the identification of the dimensions of evaluation capacity in federal government organizations. This approach, as well as other approaches that have been derived from it, are described more fully in Chapter 3. These approaches share the common elements of identifying the "ideal" state of a specific aspect of organizational growth and the current state of the organization in relation to this goal. A series of "manageable steps" are then developed to highlight the various milestones to be achieved on the way to the ideal state.

Other authors, such as Horton et al. (2003) and Lusthaus, Adrien, Anderson, Carden and Montalvan (2002), have used their own frameworks to assess organizational capacity in the context of international development. Their work highlights the value of an organizational self-assessment approach in evaluating capacity development and is similar to most participatory evaluation approaches.

The framework developed by Lusthaus et al. (2002) focuses on measuring organizational performance, understanding the organization's external environment, determining organizational motivation and examining organizational capacity. Each of these four key concepts is developed into more concrete dimensions of organizational capacity, such as strategic leadership, financial management, and infrastructure. Although useful at a conceptual level, none of the dimensions provide examples of concrete actions or behaviours at either the individual or organizational

levels. In this regard, Leithwood and Montgomery's (1987) innovation profile approach is superior to that developed by Lusthaus et al (2002).

The factors influencing evaluation capacity and the stages of ECB have been identified conceptually in the evaluation literature, as demonstrated in the previous sections. However, no study has yet identified the dimensions of evaluation capacity and describe how these would be manifested in organizations. Such information, I would argue, would advance the current state of knowledge in the field of evaluation and provide an interesting backdrop for further work in this area.

Challenges to the Implementation and Study of Evaluation Capacity

Although ECB is increasingly a contemporary focus for empirical research, several challenges related to the implementation and study of organizational evaluation capacity can be identified from the existing literature.

First, as described previously, the distinction between evaluation capacity and ECB is sometimes blurred. ECB, as defined earlier, refers to the intentional effort undertaken by an organization to become more reflective and more involved in evaluation processes, while evaluation capacity refers to the concrete practice and use of evaluation in an organization. For example, an excellent definition of ECB is provided by Baizerman et al (2002) in their summary and analysis of 11 case studies each relating ECB efforts in different organizations. A clear definition of evaluation capacity itself, however, is not included in the paper, nor does it include specific clues as to how evaluation capacity might manifest behaviourally in an organization.

Second, no clear link has been established between evaluation capacity building initiatives and organizational learning. The reflective case narrative frequently applied in ECB research has not yielded comprehensive data about the linkages that may exist between ECB and

OL (Cousins et al., 2004; Robinson & Cousins, 2004). While the reflective narrative represents a legitimate way of knowing and yields interesting suggestions in terms of the potential dimensions of evaluation capacity, it does not offer an empirically-derived conceptual framework that can be explored further or replicated. Examples of these case narratives were provided throughout the literature review (e.g., Compton et al., 2002; King, 2002) as illustrations of the state of the knowledge on ECB. Other examples following the same general direction abound, such as a case study by Porteous and Sheldrick (1999) of the development of an evaluation tool kit for Ontario public health managers. Although the article does describe how the tool kit and guiding principles for evaluation were developed, the paper focuses mostly on a description of the tool kit and offers no real insight into how evaluation capacity might be built in public health organizations.

There also remains a need for a theory of evaluation capacity at an organizational level, particularly in terms of its behavioural manifestations in organizational life (Milstein et al., 2002). Inquiry in this field, therefore, would benefit from a focus on establishing a theory of evaluation capacity. More specifically, the identification of the key dimensions of evaluation capacity as well as the manageable steps required to move from typical to high capacity levels may constitute a useful first step in better defining evaluation capacity as a contribution to research and theory. From the practical perspective, the framework through which this knowledge would be operationalized may offer organizations a model on which they can reflect on their own capacity development activities.

Last but not least, as described in the first chapter, the context within which evaluation is conducted in Canadian federal government organizations has begun to change over the last few years from a formative, or discovery-based type of evaluation, to a more summative, or

accountability-focused form of evaluation. This represents a departure from the more common results-based management espoused by governments worldwide and means that capacity building may take a back seat to audit-like studies that remove evaluation from the hands of program staff. Indeed, the responsibility for evaluation is increasingly going to external evaluators charged with making broad, summative statements about the merit and worth of a program. This also reflects a shift in mentality about potential uses of evaluation: formative evaluation typically serves the needs of program managers and staff interested in bringing about improvements to their programs, whereas summative evaluation is best suited to the informational needs of senior executives concerned with the distribution of resources within the organization. In the eyes of the latter stakeholders, evaluation could be perceived as a control mechanism as opposed to one of improvement. The value of evaluation as an enterprise of discovery, therefore, may be threatened in the context of this study. It is hoped that the framework developed as an outcome of the study will support government organizations still interested in developing their evaluation capacity by providing suggestions on potential avenues of improvement.

Research Questions

As shown in the previous section, the existing literature on ECB is largely based on case studies, many of them reflective narratives, that attempt to identify and describe the factors hypothesized to contribute to ECB success; these papers do not weigh one factor against another and, therefore, little is known about the relative contribution of each factor to overall ECB success or the stages at which each factor is most likely to have an impact on ECB. In addition to this, very little work has attempted to identify the actual dimensions of evaluation capacity and the way in which these dimensions might be manifested in organizations. The proposed study

intends to address these deficiencies in our knowledge and understanding by addressing the following research questions:

1. What are the essential dimensions of evaluation capacity in Canadian federal government organizations?
2. How are minimal and exemplary performance on each of these dimensions characterized?
3. What are the manageable steps required to move from minimal to exemplary performance?

CHAPTER 3 – METHODS

Overview of Research Design

The data collection approach used in this study consisted of three phases. The first phase sought to identify the key dimensions of evaluation capacity in Canadian federal government organizations and characterize minimal and exemplary performance on each dimension through the development of a framework. The second phase focused on refining the framework by presenting it to the interviewees from the first phase and obtaining their thoughts on the clarity of the wording used to describe the key dimensions of evaluation capacity. The third and final phase of data collection sought to triangulate the findings included in the revised version of the framework by asking individuals in four different organizations to use the framework in their own work environment and to provide feedback on whether the dimensions included in the framework enabled them to better understand their organization's current level of evaluation capacity and provided them with useful information on potential next steps. The three phases are summarized graphically in Figure 2 following the next section.

This section briefly describes the reasons behind the methodological approach selected for this study as well as the epistemological assumptions underpinning this research. A thorough description of the method followed during all three phases of the study is then provided, including a description of the participants in each phase as well as an explication of the criteria and process used to structure data collection and analysis.

Basis for Selected Approach

The Innovation Profiles approach developed by Leithwood and Montgomery (1987) features a framework for planned change of classroom practices and will be used as a basis for the framework that will be developed as a result of the identification of the dimensions of

evaluation capacity in federal government organizations. Conceptually, this approach focuses on a process of “gap reduction” – in other words, it involves identifying the “ideal” state of a specific aspect of organizational growth and the current state of the organization in relation to this goal. A series of “manageable steps” can then be developed to highlight the various milestones to be achieved on the way to the ideal state. These manageable steps should be challenging enough to lead to observable change from the previous state, but at the same time should be feasible in order to foster success in moving between steps. This approach was used as an alternative approach to process evaluation in the context of mental health case management by Cousins, Aubry, Smith Fowler, and Smith (2004). Their process involved the development of an instrument - entitled “Key Component Profile” – that was used to measure multiple dimensions of intensive case management performance. Such an instrument is presented as a matrix in which each dimension or sub-dimension of performance is described through sets of behavioural descriptions ranging in feasible steps from typical levels of performance to exemplary performance.

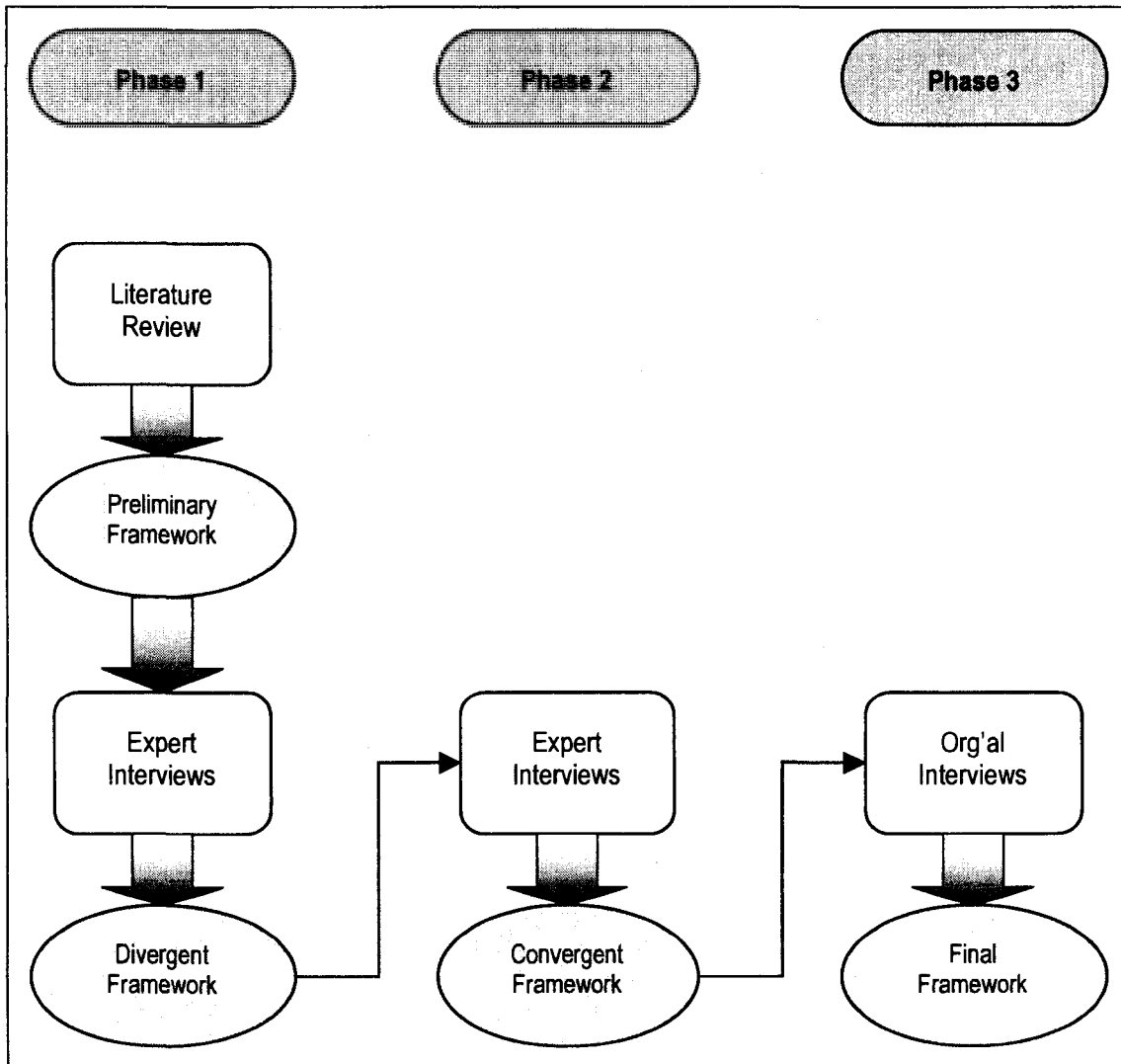
The approach developed by Leithwood and Montgomery (1987) and adapted by Cousins et al. (2004) was well-suited to a study on evaluation capacity because of its focus on the incremental steps required to move from low to high capacity. Its flexibility in terms of the inclusion of varying numbers of levels as well as an unlimited number of dimensions highlight its usefulness within the context of this study.

A key component of the approach developed by Leithwood and Montgomery (1987) is the interviewing process used to collect the information required to build the Innovation Profile. Although these authors refer to both interviews and focus groups in their examples, only interviews were used here (aside from an in-depth literature review). According to Patton (1997),

the purpose of interviewing is to “allow us to enter into the other person’s perspective” (p.341). Clearly, the main advantage of interviewing is the ability to clarify statements and to gain more insight into the perspective of the participant than would be possible with other data collection mechanisms, because the richness of the data collected through interviews far exceeds that obtained with other, more quantitative, methods. At a more technical level, interviews typically have a high response rate, and they are less demanding in terms of the reading and writing skills of the participants. In areas in which little research has been done, interviews are often a good starting point, because they can generate useful information or provide theoretical dimensions on which other data collection instruments can be based.

Interviews were used in all three phases of the study (described later in this chapter). A separate interview protocol was developed for each phase with keen attention placed on each data collection instrument’s technical features. Credibility, or dependability, can be assessed through member checks of interview data, peer debriefing, and observation (Lincoln & Guba, 1986). In the case of this study, this was achieved mainly through observation – because the interview participants were by and large experts on the object of the study, it was assumed that any obvious error would be raised by the interviewees. Transferability requires thick, descriptive data. This was also achieved by encouraging participants to respond to open-ended questions on the dimensions of evaluation capacity and other aspects of organizational life related to this topic (Creswell, 1998; Lincoln & Guba, 1986). Both of these technical features increase the trustworthiness and authenticity of the data collected and, thus, ensure that these are representative of the participants’ experiences with the object of the study (Creswell, 1998).

Figure 2: Summary of Methodology



Epistemological Assumptions

The process of reviewing the literature on this topic and the interviews that took place during the study were grounded in a logical empiricist approach. In other words, the use of existing, preordained knowledge on evaluation capacity in the identification of the dimensions and the intent to use the framework in a transferable manner across contexts assumed a logical empiricist stance, even though the consultative methods used to further refine and validate the dimensions were reflective of social constructivist principles and practices. The implementation

of the framework describing the dimensions across different contexts will depend on local interpretation and meaning making, but given the scope of this project, the main theoretical approach remains that of logical empiricism. This theoretical approach borrows from Greene's (2002) mixed-method approach to evaluative inquiry and is reflective of Huberman's "sustained interactivity" model (1994), both of which call for the involvement of various stakeholders throughout the research process in an attempt to foster knowledge use. The pragmatic stance adopted by such authors supports a compatibilist view of the positivist and interpretivist paradigms and allows for the integration of quantitative and qualitative methods at the levels of data, design and analysis, and interpretation (Howe, 1988; Phillips, 1990).

Although only qualitative data were used in this study, the overriding concern to produce a framework that would be of use to practitioners while contributing to theory on evaluation capacity indicates a need for an epistemological standpoint imbued with social constructivism and post-positivism. The study attempts to draw on shared understandings of what evaluation capacity looks like in government organizations (social construction of knowledge through sharing of experiences and thoughts during the interview process) while acknowledging that the dimensions of evaluation capacity, although intangible, can be identified in an objective manner beyond the direct experience and environment of the participants. This results in a series of findings that are transferable across similar organizations that may not have been involved in the study.

Role of Researcher

All three phases of the study described in the next few sections were conducted by the student researcher under the guidance of her thesis supervisor. This involved conducting the in-depth literature review, identifying and communicating with potential interview participants,

conducting and recording the interviews, transcribing the interview data, conducting the data analysis, and writing up the findings. The impact of the student researcher's unique position as both a researcher and an evaluation officer in a Canadian government department varied depending on the individual interviewed. For instance, as noted in the section describing the participants for the first phase of the study, two of the four evaluation experts interviewed already had a professional relationship with the student. In this case, the interviews were fairly informal and took on more of a conversational, collegial flavour. One of the other two interviewees was known to the student and knew of her, so their first meeting was also very collegial and conversational. The fourth participant, however, took on a more formal role in the interview, partly because of his role within the Canadian evaluation community, but also because he seemed very aware of his participation in a research project. Clearly, the presence of the student researcher had an effect on the comments made by participants in the first phase: while in three of the four cases, the student was able to create a comfortable, open climate, this was more difficult to achieve in the fourth case. Similar observations can be made in the third phase of the study, when heads of evaluation, senior evaluation officers, and program managers in four departments or agencies were interviewed. While the heads of evaluation and senior evaluation officers treated the student much as a colleague or peer, the program managers were less clear about the student's role and asked more questions about how the data would be used in the study. Generally, the heads of evaluation were more direct in both the feedback offered on the framework and their organizational assessment, whether positive or negative, while the senior evaluation officers often tempered or massaged their comments, especially in terms of organizational assessment. A notable exception to this is the case of the organization with low capacity: the head of evaluation was new to the position and had no evaluation background,

while the senior evaluation officer had considerable experience and credentials in evaluation. In this case, the roles seemed to be reversed. Further details on the organizational roles of participants in each phase of the study will be provided in the next few sections of this chapter. A summary of the participants by phase of the study is provided in Table 1 below.

Table 1: Participants by Phase of Study

Participant Type	Phase 1	Phase 2	Phase 3
Experts	<ul style="list-style-type: none"> ▪ Four participants with a broad view of evaluation in government ▪ Organizational roles involve consulting and directing the government evaluation function from a central agency 	<ul style="list-style-type: none"> ▪ Same participants as Phase 1 	
Heads of Evaluation			<ul style="list-style-type: none"> ▪ Four Heads of Evaluation at the Director or Director-General level
Senior Evaluators			<ul style="list-style-type: none"> ▪ Four senior evaluators with considerable experience and expertise
Decision-Makers			<ul style="list-style-type: none"> ▪ One program manager, one evaluation manager situated within a program and one senior executive ▪ Able to comment on evaluation use in their organizations

Phase 1: Identification of Key Dimensions of Evaluation Capacity (Divergent Phase)

The Innovation Profiles method developed by Leithwood and Montgomery (1987) was used as a basis for the identification of the dimensions of evaluation capacity in federal government organizations. The first phase of the study focused on the development of a framework through which the key dimensions of evaluation capacity were operationalized. The

framework was developed using two data sources: an in-depth literature review and a set of semi-structured interviews. Each of these will be described in turn.

In-depth Literature Review: Procedures and Data Analysis

The goal of the in-depth literature review was to refine the literature review conducted in the development of the research questions in order to identify the key dimensions of evaluation capacity and examples of minimal and exemplary performance for each dimension. The process used to accomplish this task was to update the literature search in six different databases (ERIC, E-Journals @ Scholar's Portal, Sage Management and Organization Studies Full-Text, Public Administration Abstracts, Social Sciences Citations Index, and Sage Education Full-Text). Key search terms used included "evaluation capacity", "evaluation utilization", "evaluation use", "capacity building", "organizational effectiveness", "performance measurement literacy", and "evaluation culture", as well as combinations of some of these terms. Once new papers on ECB were identified, these were reviewed and summarized on one-page documents highlighting key points. The reading notes produced were then integrated into the existing literature review. An important aspect of the literature review in this phase of the study involved moving beyond descriptions of capacity building initiatives undertaken in various organizations and focusing instead on the definitions and features of evaluation capacity itself and how it appears in organizations. The general guide used to identify such instances in the literature was the definition of evaluation capacity provided by Boyle, Lemaire, and Rist (1999): "the human capital and financial or material resources required to conduct evaluation". Clues to possible dimensions were highlighted in the reviewed texts and a preliminary framework summarizing these was developed in order to inform the subsequent phases of the study. The results of this phase are presented in Chapter 4. Once the literature review was completed, semi-structured

interviews were conducted with expert informants working in the field of program evaluation and who have a broad view of evaluation in the Canadian federal government. These interviews sought to collect data on the experience of these participants in order to ensure that the context within which the framework was developed (i.e., federal government) was taken into account.

Semi-Structured Interviews: Participants

Four individuals were asked to participate in the first phase of the study; two were external consultants who have worked with several departments and agencies on evaluation studies and who therefore have experiential knowledge concerning the evaluation practices and capacity of different departments and agencies; two were former or current senior officials of a central agency of the government of Canada who have worked on interdepartmental evaluation issues and who are familiar with the challenges faced by different departments and agencies as they develop their evaluation capacity. These individuals are part of the researcher's community of practice and therefore already knew either the researcher or her supervisor; two of these individuals had already expressed interest in participating in the study. The individuals were selected by the student researcher, in conjunction with her supervisor, based on their experience, expertise, and position relative to the federal evaluation function. Their point of view as insiders of the federal evaluation community but outsiders of the evaluation function of specific departments and agencies is key to their overall vision of how evaluation capacity appears in various organizations. In-depth contextual and biographical information on each expert participant is provided in the next few paragraphs.

Expert 1 is a private-sector consultant specializing in management consulting, performance measurement, and evaluation. This individual has over 25 years of experience in this role. He was selected for the study because of his extensive experience with Canadian

federal government clients and because of his considerable contributions to the Canadian Evaluation Society and the American Evaluation Association. He was also heavily involved in the development of the requirements for the Results-based Management and Accountability Framework (RMAF) introduced by the Treasury Board Secretariat in 2001. This participant was chosen because of his keen interest in developing the evaluation capacity of organizations and his vantage point as an observer of federal government evaluation practices and policies. The student researcher and her thesis supervisor had established a professional relationship with this individual through various activities prior to the start of the thesis research; this made him a natural choice as an expert participant.

Expert 2 has over thirty years of experience in evaluation and performance measurement, having managed evaluations in a variety of public sector settings and leading corporate evaluation units in five Canadian government departments and agencies. Most recently, this individual has been directly involved in evaluation policy and implementation within a federal government central agency. He also played an instrumental role in the creation of the Results-based Management and Accountability Framework (RMAF), and was responsible for rebuilding and repositioning the evaluation function within the Government of Canada. Evaluation capacity building has been of interest to him throughout his career. This participant was selected because of his extensive knowledge of the evaluation function throughout government. The student researcher and her thesis supervisor had also established a professional relationship with this individual through various activities prior to the start of the thesis research.

Expert 3 has also worked as an evaluator for several years and is well recognized among peers for her contributions to the field of evaluation. Her work experience spans both the public and private sectors. In addition to her work as an evaluator, this individual has been closely

involved with the Canadian Evaluation Society and has taught evaluation and research methods at the university level and has lead a number of professional development workshops. This participant was selected because of her considerable experience with federal government clients and her commitment to building evaluation capacity in her own work environment. This participant was also known to the student researcher, although they had not met formally, through mutual acquaintances.

Expert 4 has more than 15 years of experience in the Canadian public service. His career path includes work in the area of policy development as well as evaluation and performance measurement, with a number of years spent in a federal government central agency. More recently, this individual was directly involved in evaluation policy development and implementation. The current proximity of this individual to the federal government evaluation community as well as his differing views from those of other participants about the role of evaluation in government made him a good choice for the study. Although the student researcher did not know this participant prior to the study, her thesis supervisor had numerous occasions to meet with this individual and was able to introduce him to the student researcher.

Although the interviews were conducted individually, the fact that the informants are part of the same community of practice introduced the possibility that their points of view would be influenced to some extent by the current evaluation philosophies and practices espoused by the federal government. This could have resulted in the development of a framework that reflects only the prevailing views about the evaluation of government programs. This risk was mitigated by using the literature review as the basis for the framework and by selecting individuals with different roles for the interviews (i.e., external consultants as well as government insiders). In addition to this, the interviews conducted with evaluators and users from different departments

and agencies in the third phase of the study provided another mechanism for encouraging diversity in the opinions and views of participants.

Semi-Structured Interviews: Procedures

The key informants described above were first contacted by the thesis supervisor of the student researcher via email and asked if they were interested in participating in a face-to-face interview that would last approximately one hour (Appendix A). Once each person had agreed, the student researcher then contacted them individually, again via email, to schedule the interview. The informed letter of consent was attached to this email (Appendix B). Three of the interviews were held at the workplace of participants and another one was done by telephone due to scheduling issues. Each interview was recorded on audiotape and transcribed. A paper copy of the letter of consent was provided to each participant at the interview for their signatures. No issues were raised by participants over the contents of the letter of informed consent or over the interview questions. Two of the four interviews lasted approximately one hour, and two lasted approximately 90 minutes. All four participants agreed to have the interviews recorded. The researcher took notes during the interview as well in case of technical difficulties related to the recording, and to add personal observations about the participant or to flag certain comments for review.

Semi-Structured Interviews: Data Collection

The interview protocol (Appendix C) was developed by the student researcher in order to ensure uniformity in the questions asked to participants and consistency in the data collected. The primary goal of the interview was to identify the key dimensions of evaluation capacity. The first question focused on setting the stage for the study and the remainder of the interview and asked participants to define evaluation capacity. This question also helped ensure that the

researcher and participants understood each others' perceptions of evaluation capacity and what this represents to prevent confusion throughout the interview. The following questions were designed to help participants reflect on evaluation capacity by situating them in an organizational setting and asking them questions based on the concrete manifestations of evaluation capacity (e.g., Can you give me some examples of Canadian federal government departments or agencies that you think reflect a high or exemplary level of evaluation capacity? What specific characteristics do these organizations have that make them highly capable in terms of evaluation practice?).

The key criteria used to guide data collection, as evidenced in part by the interview protocol, but also more informally through prompts by the student researcher during the interviews, were breadth/comprehensiveness (e.g., What would you look for in an organization in order to assess its evaluation capacity), clarity (e.g., In what sense is their capacity lower than other organizations?), manageability (e.g., In your opinion, is this something that is important to government organizations? Has it worked?), and utility (e.g., Do you know of any organization that has clearly benefited from building its evaluation capacity, government or otherwise?).

In order to test the interview protocol prior to the start of data collection, the student researcher asked colleagues to review the instrument and comment on its technical properties. Two professional evaluators were asked to read over the protocol and provide feedback related to the clarity of the questions asked as well as the contents and comprehensiveness of the instrument. Minor changes were made to the instrument as a result of the feedback obtained. These changes were related to wording and did not affect the overall sense of the questions included in the protocol.

Semi-Structured Interviews: Data Analysis and Reporting

A qualitative content analysis was used to identify trends in the interview data. Because of the small number of participants, the volume of data generated was manageable and specialized qualitative analysis software was not required.

Each participant was the object of an individual data file, with the interview protocol used as template for the transcriptions. The handwritten notes were transcribed first into the template for each participant and were supplemented by the recordings. Special notes taken by the researcher on context, or to flag certain pieces of information, were written in a separate colour to make them more visible during the analysis and quotations to be used in the thesis were highlighted. Once all of the interviews were transcribed, the responses of each participant were colour-coded for ready identification and a master file was created to include all of the participants' responses to each question. The researcher then reviewed the data for each question individually and highlighted all of the nuggets of data relating to the dimensions of evaluation capacity. These pieces of information were grouped according to their general meaning, and a common name was derived for each group of data. For instance, this process yielded dimensions such as "Human Resources" and "Use of Consultants". At this stage of the research process, all pieces of information were considered relevant, whether only one participant had identified a particular dimension of evaluation capacity or all participants had named it during the interviews. An additional data category was created to include data that did not specifically refer to a dimension of evaluation capacity, but instead focused on the manifestation of evaluation capacity in an organization. This category was named "markers of evaluation capacity" and contained information such as "increased resources to the organization" and "demand for evaluation

services". These data were later used to describe certain dimensions or sub-dimensions of evaluation capacity in further detail.

The second stage of coding involved the identification of government organizations at high, intermediate and low levels of evaluation capacity. Once again, each question was analyzed individually and a list of organizations was derived based on the comments of participants. The specific comments of participants pertaining to each of the organizations were carefully derived from the interview data. The number of participants suggesting each of the organizations was also noted in the data file in order to base the selection of organizations for Phase 3 of the study on the opinion of more than one participant whenever possible.

The interview data were used to complete, complement and challenge the current literature on evaluation capacity. The data were carefully compared against the existing framework based on the literature review, and the guiding principle for the analysis was the required changes to the preliminary framework based on the interview data. The process used to update the preliminary framework as well as the specific changes made are detailed in Chapter 4 of the thesis.

Phase 2: Review and Feedback on Draft Framework (Convergent Phase)

The second phase of data collection focused on confirming the key dimensions of evaluation capacity derived from Phase 1 activities. This phase once again involved key informant interviews with those individuals interviewed in the first phase of the study. At this point, participants were asked to review the draft framework and to provide feedback on its clarity and contents.

Procedures

The four participants from Phase 1 were contacted by the student researcher via email after the completion of a revised framework of evaluation capacity summarizing the results of the first phase of the study. All participants had agreed to a second interview approximately one month after the first interview. The informal email sent out by the student researcher included the revised framework and highlighted to participants that this would be the object of the interview. All four interviews were scheduled within a few weeks of this email. Three of the four interviews were held at the participants' office, while a fourth interview was held in a public location due to lack of meeting space. This participant selected a convenient place within walking distance of his home. As for the first set of interviews, all four interviews were recorded. However, the recording taken during the fourth interview was unusable due to the amount of noise recorded in the public location used. In all cases, notes were taken during the interviews, so the impact of not recording the fourth interview was minimized in terms of the subsequent data analysis. The informed letter of consent signed by participants in Phase 1 also included their participation in Phase 2, and so no further consent was required at this stage of the study. All four interviews lasted between 60 and 90 minutes.

Data Collection

The interview protocol (Appendix D) was once again developed by the student researcher in order to ensure uniformity and consistency throughout the interviews. The main purpose of this second interview was to obtain feedback on the revised framework of evaluation capacity. The first question sought to confirm that the participant had indeed reviewed the framework sent via email, and to give a chance to the participant to read it over if this had not already been done. All four participants had taken the time to read the framework before the interview, although two

reviewed it quickly to refresh their memories. The second question focused on the breadth and comprehensiveness of the dimensions included in the framework (e.g., Does the framework include the most important dimensions of evaluation capacity? Is anything missing?). The third question also focused on the framework, but sought to obtain information about the clarity of the wording used to describe each of the dimensions (e.g., Is the wording used for the different column headings clear?). The fourth question sought to collect data on the actual steps required to move from a low capacity level to a high capacity level and to determine if these were described in sufficient detail (e.g., I'd like you to tell me if you think the steps are manageable, and if they are clearly outlined in the framework). The following question dealt with the potential usefulness of the framework in an organizational setting, to ensure that no outstanding issue had been left out of the conversation (e.g., Do you think that this framework could be used by organizations interested in self-assessing their level of capacity?). Finally, the last question required the participant to suggest organizations that would fit within the two extremes of the framework in terms of capacity, in order to refine the selection of organizations for the third phase of the study (e.g., Can you recommend organizations that have a typical level of evaluation capacity and others that have an ideal level of evaluation capacity?). This last question was not used during the actual interviews, because the student researcher felt that sufficient time had been spent discussing this issue during the first interview and was concerned about repetition and taking up the valuable time of the expert participants.

Two evaluation professionals (the same individuals as for the first phase) were asked for their impressions of the instrument as a proxy for a pilot test. No specific comments were made by these individuals.

Data Analysis and Reporting

As for the first phase, a qualitative content analysis was used to identify trends in the interview data. The same process for transcribing and coding the data was used (i.e., creating an individual data file for each participant based on the interview protocol). The same participant coding structure was used to ensure consistency and quick referencing. Once again, a master file was created containing all of the participants' responses to all of the questions included in the protocol, with quotations highlighted in a separate colour.

The data codes used in this phase of the study focused on specific changes to be made to the framework. The data were divided into two main categories, one that included all comments made about the framework at a general level, and one that included all of the specific comments made on each of the dimensions. Each of these categories were further refined. The category outlining the general comments about the framework was divided into "name and number of levels used", "applicability of framework", "format of framework", and "contents of framework – general". This last category contained information related to additional references suggested by participants, statements about how the framework's dimensions could be measured, and comparisons with the new evaluation policy. The second main category was divided up according to the structure of the framework and included comments on the wording of the dimensions as well as the contents included within each one, organized by sub-dimension. A final category, called "additional elements", was also created to include nuggets of information that did not fit within the other categories. This included participants' thoughts on the role of central agencies in increasing the capacity of departments and agencies to evaluate and use evaluation results, as well as the need for evaluations to cut across themes within the organization rather than focus on a small number of programs.

The coded and organized interview data were interpreted as confirmations of existing dimensions and sub-dimensions or as challenges that would warrant some changes to the information included in the framework. The process used to modify the framework based on the Phase 2 data as well as the specific changes made are detailed in Chapter 5 of the thesis.

Phase 3: Triangulation of Findings Included in the Framework

A validation exercise was undertaken in order to finalize the draft framework of evaluation capacity. This exercise was conducted to triangulate the data obtained through the first and second phases of data collection. This phase therefore focused on key informant interviews with evaluators and decision-makers from selected federal government departments and agencies. These individuals were asked to use the framework in their own organizations and provide feedback on its potential utility in terms of organizational reflection and improvement.

Organizations Selected for the Validation Exercise

Five organizations were selected by the student researcher and her thesis supervisor based on the Phase 1 interview data. It was felt that conducting interviews in five organizations would be manageable enough for one researcher and would provide sufficient data to validate the framework, as long as there was some variability in the evaluation capacity of the organizations studied. Two organizations at the exemplary level (Organizations 1 and 2), one organization at the intermediate level (Organization 3), and two organizations (Organizations 4 and 5) at the low level were identified to ensure such variability. The Heads of Evaluation of the five organizations were contacted by the thesis supervisor in an informal email and asked if their organization would be interested in participating in the study. Four of the five organizations agreed to participate, but Organization 5 was unable to contribute to the study due to changeover in leadership and a lack of ability to find the time and space necessary to engage in reflection

about evaluation capacity. Further information about each of the four participating organizations is included in the following paragraphs.

Organization 1: This large service organization (over 20,000 employees across Canada) was identified as exemplary by the Phase 1 participants because of its investment in technical capacity and the rigour of its evaluation products. It has a large team of evaluators, divided up into five units organized by program area. This way, evaluators in each team become experts in a particular subject area and are able to gain a deep understanding of a small number of programs.

This organization has received numerous awards for the quality of its evaluation work and leads the evaluation community by example. Some of the more innovative aspects of the work conducted in this department include the extensive use of peer review as a quality control mechanism and the development of complex quantitative evaluation methodologies.

Organization 2: This medium-sized regulatory organization (approximately 4,000 employees) had also been identified by the Phase 1 participants as having exemplary capacity because of its evaluation unit's continued efforts to reach out to program areas and to diversify the products offered to its clients. Like the other exemplary-level organization studied, this organization has also received awards for its efforts and is often held up as an example to follow for other government departments and agencies.

Organization 3: This medium-sized organization (approximately 4 000 employees) had been identified as being a strong intermediate by the Phase 1 experts. These participants recognized recent efforts made by the Head of Evaluation to build the organization's evaluation capacity on several fronts, such as staffing, professional development, ongoing data collection, and management processes. It was selected because of these reasons, and also because its unique mandate within the area of science and technology raise specific evaluation challenges not found

in other types of government organizations (e.g., Evaluating programs of research vs. evaluating research itself).

Organization 4: This organization has an operational mandate within the federal government and has therefore only recently been required to evaluate its programs. Because it is still in the early stages of developing a central evaluation function, its capacity level was found to be low by the Phase 1 participants, who cited a high turnover rate and a low skill level as the main reasons behind their assessment.

Participants

The student researcher sent an official invitation to the Head of Evaluation of each of the four participating organizations to provide further details about the study and what organizational participation would imply (Appendix E). Attached to this email were an individual letter of informed consent (Appendix F), an organizational letter of informed consent (Appendix H) and the revised framework summarizing the Phase 2 findings. The student researcher then scheduled the interview to be held with each Head of Evaluation via email or telephone, depending on each participant's preference. These participants were also asked to nominate two other members of their organization for participation in the study (one senior evaluation officer and one organizational decision-maker). The Heads of Evaluation contacted the other two participants from their organization and provided them with the student researcher's contact information, which they could use if they decided to participate in the study. In all cases but one, the additional participants contacted the student researcher and the interviews were scheduled within that month. One individual from the low capacity organization could not be reached and no suitable replacement could be identified by this organization's Head of Evaluation to participate in the study. Therefore, one organization is represented by only two participants in the study;

further details on this are provided below. The individual letter of informed consent was sent out to all participants via email along with the revised framework.

Although in-depth biographical information was not requested of participants to maintain their confidentiality, some information about their levels and experience is known. This information is related here to provide more context to the individual participants' responses to the interview questions. Codes used throughout the thesis to identify the participants are provided in brackets after each description.

Organization 1 – Head of Evaluation: This individual is the Director General of evaluation services and was appointed to this position following a long career as departmental evaluator and director of one of the evaluation groups within the organization (Head of Evaluation 1).

Organization 1 – Senior Evaluator: This individual is the Director of one of the evaluation teams within the department and has worked as an evaluator for two different government departments since the completion of a Ph.D. degree in sociology (Evaluator 1).

Organization 1 – Decision-Maker: This individual is the Director of one of the key programs managed by the department. His area of substantive expertise lies in policy development, but he often requires evaluation work and is very familiar with the processes and procedures in place to ensure quality evaluation work within the department (Decision-Maker 1).

Organization 2 – Head of Evaluation: This individual has been Director of the evaluation function of the department for a number of years. Under her direction, the department has adopted a results-based orientation and the services provided by evaluators have expanded to include facilitation and other activities related to performance management. This participant is often considered to be the voice of government evaluators in the broader evaluation community,

and has assumed a role of leadership both within and outside of her organization (Head of Evaluation 2).

Organization 2 – Senior Evaluator: This evaluation professional has been with the department for a number of years and has considerable experience conducting evaluation studies and managing contracts related to evaluation work. Her most recent challenge lies in managing a recruitment drive for senior evaluators on behalf of a number of departments and agencies; this work has lead her to develop substantive knowledge about evaluation capacity building from a staffing perspective (Evaluator 2).

Organization 2 – Decision-Maker: A program specialist responsible for evaluation was selected by the organization’s head of evaluation as a potential participant due to his recent exposure to a large-scale evaluation project. This individual works directly within a program group and manages evaluation-related projects. However, this person’s definition of evaluation is somewhat different than the generally accepted notions of evaluation held across government. He is responsible for the development of “evaluation protocols”, a set of guidelines used in the assessment of applications made to the department for funding. In other words, he provides assistance towards program delivery and not program evaluation. Although this person did not fit within the usual role of decision-makers involved with evaluation, the judgment of the department’s head of evaluation was respected and this participant was included in the study. More details about this person’s views on evaluation and their impact on the data collected are provided in Chapter 6 (Decision-Maker 2).

Organization 3 – Head of Evaluation: This individual heads a group of professionals working in the areas of evaluation, performance management and planning. Prior to joining this particular organization, the director of this group worked for a number of years as a consultant

and has considerable experience at every step of the evaluation process. She has been involved with the Canadian Evaluation Society for a number of years and has lead a number of initiatives on behalf of its local chapter (Head of Evaluation 3).

Organization 3 – Senior Evaluator: This individual also has a number of years of experience as an evaluator, both in the public and private sectors. She has managed several complex, large-scale evaluation projects and also handles the day-to-day management of the evaluation team for her department (Evaluator 3).

Organization 3 – Decision-Maker: The decision-maker selected for this organization sits at a Vice-President level and is a member of the organization's Senior Executive Committee. His thoughts on evaluation practice and capacity within his organization reflect his experience with corporate matters and the visioning process that is a key feature of management in this organization (Decision-Maker 3).

Organization 4 – Head of Evaluation: The Director of evaluation of this organization had only been in this position for a few months at the time of the interview. This person had a background rooted in policy work and had only been exposed to evaluation in a general manner, with no formal training or direct work experience in this area (Head of Evaluation 4).

Organization 4 – Senior Evaluator: This individual had considerable experience in the area of evaluation and performance management, obtained in a number of different departments and agencies. Although evaluation is fairly new to this organization, this individual had clearly reflected on how best to build evaluation capacity from a human resources perspective and from an organizational standpoint (Evaluator 4).

Data Collection

The purpose of this phase of the study was to obtain validation of the dimensions and sub-dimensions included in the framework by asking individuals in four organizations to use the framework and to provide feedback on the contents and wording used in the framework and on its applicability to their particular situations.

The interview protocol (Appendix G) was developed by the student researcher in order to ensure uniformity and consistency throughout the interviews. The first question sought to confirm that the participant had indeed reviewed the framework sent via email, and to give a chance to the participant to read it over if this had not already been done. Most participants had taken the time to read the framework before the interview, although three stated that they would like to review it to remind themselves of its contents. Along the same lines, the second question sought to identify whether the participant had consulted others in conducting the organizational assessment, to get a better sense of the number of people involved in using the frameworks. None of the participants involved others in using the framework, and so no further questions were asked on this issue. The third question was the same as one of the questions asked in the second phase of the study and focused on the breadth and comprehensiveness of the dimensions included in the framework (e.g., Does the framework include the most important dimensions of evaluation capacity? Is anything missing?). Similarly, the fourth question focused on the clarity of the wording used to describe each of the dimensions (e.g., Is the wording used for the different column headings clear?). The fifth question sought to collect data on the actual steps required to move from a low capacity level to a high capacity level and to determine if these were described in sufficient detail (e.g., I'd like you to tell me if you think the steps are manageable, and if they are clearly outlined in the framework). Finally, the last question required the participant to

discuss their organization's evaluation capacity, based on the dimensions included in the framework (e.g., Can you tell me about your organization's capacity, based on the framework?) and provide feedback about the ease of use of the framework to accomplish this task (e.g., How difficult was it to use the framework?).

Because many of the questions included in this protocol were the same as those of the second phase protocol, the data collection instrument was not pilot tested. The second phase of data collection was considered an appropriate proxy for piloting the instrument, since no major issues related to the qualities of the protocol were raised in the other segments of the study.

Data Analysis and Reporting

As with the interviews conducted in the first two phases of the study, a qualitative content analysis was used to identify trends in the interview data. Because of the increased complexity associated with the use of four different organizations and three different organizational roles, the data coding and analysis processes were more detailed than for the first two phases.

Each interview was first transcribed based on the interview protocol as had been done previously. In this case, two participants refused to have their interviews transcribed, so recorded data were only available for nine of the eleven interviews. The two other interviews were carefully transcribed based on the notes taken by hand.

Data from the individual interviews were first grouped by organization, with separate colour coding to track each participant's contribution to the combined organizational data. Individual pieces of information were then divided into comments about the framework and comments about the organization in order to clearly distinguish between the feedback provided on the framework and the organizational assessment that took place subsequently. The data focusing on the framework itself were then categorized by sub-dimension, with more general comments

kept separate from the rest of the data. Data from all participants on the framework for each of the sub-dimensions were then grouped into one master file; within each sub-dimension, data were further coded based on whether they referred to wording issues, problems related to differentiating between the levels, or suggested additions and comments. These data were then used to identify the changes to be made to improve the framework. At this stage, the weight of one type of participant versus another (i.e., Head of evaluation compared to evaluation user) was not considered in the analysis – because the analysis focused on improving the framework, all pieces of information were treated equally by the student researcher.

Data obtained on organizational capacity, or the assessment made by each participant using the framework, were also categorized based on sub-dimensions of the framework. In this phase of the analysis, each organization was considered separate from the others and so four master files were created. The data included in each master file were colour coded to ensure traceability back to each individual participant. The organizational role of interviewees was preserved in the data to properly interpret the combined organizational assessment: while Heads of evaluation had more authority in terms of assessing sub-dimensions such as staffing, the evaluation users would have more exposure to decision-making processes. The consideration of organizational roles was maintained throughout the analysis. The coded and analyzed data were used to develop organizational profiles of the four participating departments and agencies.

The final version of the framework as well as the organizational profiles generated from the findings are located in Chapter 6 of the thesis.

Chapter 4 – Findings from Phase 1

The first phase of the study involved an in-depth review of the literature on evaluation capacity building as well as a series of interviews with evaluation experts who have regular contact with Canadian federal government organizations in the context of their work. The purpose of the first phase of the study was to identify the key dimensions of evaluation capacity in these organizations. The main outcome of this phase is a draft framework outlining these key dimensions according to Leithwood and Montgomery's (1987) Innovation Profile methodology.

The first section of this chapter summarizes the findings obtained through an in-depth review of the literature. These findings were used to develop a preliminary framework of evaluation capacity, presented in Table 1. The second section describes the results of the interviews that took place during this phase of the study. These interviews provided the data needed to revise the preliminary framework and are summarized in Table 2. The two last sections of this chapter present interview data pertaining to the general markers of evaluation capacity and to the capacity of specific Canadian federal government departments and agencies.

In-Depth Review of the Literature

The first step of this phase was an in-depth review of the literature focusing on a preliminary identification of the key dimensions of evaluation capacity. This step differed from the original review of the literature conducted as background to this research project in that its only purpose was the identification of possible dimensions of evaluation capacity. Publications studied as part of the original literature review were once again studied using this new lens, and a tentative list of key dimensions and how these may be manifest in organizations was drafted and organized into the framework presented in Table 1.

The key dimensions identified were typically presented in the literature as they would be observed in organizations with exemplary evaluation capacity. This level was therefore used as the starting point in developing the framework. In other words, the description of an exemplary level was first elaborated for each of the dimension categories identified in the literature review, based on the content of the publications reviewed. A description of the typical (or low) level of capacity for each dimension was then developed as an opposite of the exemplary capacity level. The preliminary framework below therefore presents two levels for all five categories found in the literature. It was expected that both the number of categories and levels would change as a result of the subsequent phases of the study, and that not all dimensions would necessarily have the same number of levels due to the varying complexity and depth of each dimension compared to the others. Nonetheless, the adoption of a simple, uniform structure for the preliminary framework made its development more systematic and provided needed clarity to this first draft.

Table 2: Preliminary Framework Based on Literature Review

	Capacity to do		Capacity to use		
	Resources for evaluation	Evaluation planning and activities	Assessment of ECB efforts	Integration with organizational decision making	Evaluation literacy of organizational members
Exemplary (High) Capacity	<ul style="list-style-type: none"> Evaluation unit adequately staffed Balance of skilled, experienced and new evaluators Evaluator roles are clearly defined Evaluation budget adequate to complete activities outlined in evaluation plan 	<ul style="list-style-type: none"> Existence of organization-wide evaluation plan In-house evaluation whenever possible Performance measurement system feeds into evaluation activities 	<ul style="list-style-type: none"> ECB implementation plan is followed Existence of a baseline of evaluation capacity to measure the effectiveness of ECB efforts Formal or informal processes to discuss lessons learned during evaluation are in place 	<ul style="list-style-type: none"> Evaluators included in management processes Evaluators in constant communication with program managers Evaluation findings considered in budget allocation and other organizational decisions Needs assessment is an integral part of program design Location of evaluation function appropriate given size and mandate of organization 	<ul style="list-style-type: none"> Organizational members are given time to participate in evaluation-related activities Performance appraisals consider participation in evaluation activities Organizational members understand the benefits of evaluation and its principles Evaluators assess their own work (peer review, etc) to model evaluation practices
Typical (Low) Capacity	<ul style="list-style-type: none"> Evaluation unit staffed mainly by junior evaluators Vacant positions within the evaluation unit Evaluator roles not defined Evaluation budget insufficient to complete activities planned for the year 	<ul style="list-style-type: none"> Independent projects conducted by external consultants in response to external demand No long-term evaluation plan for entire organization Performance measurement system does not exist 	<ul style="list-style-type: none"> Evaluation capacity building is not part of the organization's priorities Evaluation projects are not discussed at the organizational level once completed 	<ul style="list-style-type: none"> Evaluation unit operates separately from program units and is not involved in management processes Little consideration of evaluation findings in organizational decision-making No needs assessment is conducted during program planning 	<ul style="list-style-type: none"> No involvement from program staff and managers due to lack of interest or time No consideration of evaluation activities in performance appraisals Organizational members are not familiar with evaluation Evaluation work is never assessed or reviewed as a model to other organizational members

Description of Dimensions Included in Preliminary Framework

Five dimensions of evaluation capacity were identified through the literature and used in the preliminary framework presented above. The dimensions themselves can be divided into two broader categories defined as “capacity to do” evaluation and “capacity to use” evaluation. These reflect the different issues that have been identified regarding an evaluation unit’s capacity versus the capacity of the entire organization. The framework highlights the key elements of each dimension as organizations move from typical (low) to exemplary (high) capacity. These dimensions are described in the following paragraphs in an effort to provide further detail than could be included in the framework.

Resources for Evaluation

Appropriate resources must be allocated to an organization’s evaluation function in order for its evaluation capacity to evolve. Organizations with a typical (low) level of capacity for this dimension do not have fully staffed evaluation units or have a high proportion of junior evaluators on staff, do not have clearly defined roles for evaluators and do not have the necessary financial resources to complete planned evaluation activities.

Organizations with exemplary (high) evaluation capacity feature adequately staffed evaluation units (Boyle et al., 1999; Gilliam et al., 2003; Milstein et al., 2002; Trevisan, 2002), with a good balance of experienced and new evaluators who work together on various parts of evaluation projects (Brazil, 1999). The role of the evaluator within the organization is defined properly and the requisite financial resources are in place to complete the planned evaluation activities (Boyle et al., 1999; Brazil, 1999; Gilliam, 2003; Kegeles, 2005; Trevisan, 2002).

Evaluation Planning and Activities

This dimension focuses on the integration of evaluation in the ongoing activities of the organization. A typical (low) level of this dimension features independent evaluation projects conducted by external contractors, usually as a result of a demand by external stakeholders or because of an internal need for instrumental or symbolic information about a program. Data collection for these evaluation projects is undertaken based on the specific questions included in the evaluation framework and does not include existing data on program performance; indeed, the organization does not possess an integrated performance measurement system on which evaluators can rely for ongoing data collection and analysis but may collect certain types of performance data at the program level. Finally, organizations at a typical capacity level do not have a long-term plan for evaluation based on risk and informational requirements of program managers and staff.

Organizations that have achieved an exemplary level of capacity on this dimension have articulated a long-term evaluation plan that takes informational needs and program decision-making into account (Compton et al., 2002). This plan both reduces the overlap in evaluation projects conducted and takes advantage of various data collection opportunities. Evaluations are conducted either in part or completely by internal evaluators who have in-depth knowledge about each program's specific context and how it fits within the greater context of the organization (Sonnichsen, 1999). Finally, evaluations conducted in these organizations also take advantage of established performance measurement systems that ensure on-going data collection and constant feedback on program implementation (Boyle et al., 1999; Brazil, 1999; Trevisan, 2002).

Assessment of ECB Efforts

Because the development of an organization's evaluation capacity is a considerable undertaking, it too requires an action plan that must be revisited and adjusted on a regular basis. The assessment of ECB efforts is therefore a key component of any organization's evaluation capacity. Organizations do not typically engage in this type of assessment. In fact, many organizations with a typical (low) level of evaluation capacity have no intention of developing their capacity and this topic is not part of the discussions held by senior managers about the direction of the organization. In fact, in these organizations, evaluations are rarely discussed once they have been completed. Organizations with an exemplary (high) level of evaluation capacity, however, have made it a priority and have articulated a clear plan to attain their objective (King, 2002; Newcomer, 2004; Trevisan, 2002). Lessons learned are discussed on a regular basis, and the organization has established a baseline on which to judge future ECB efforts (Bozzo, 2002; Cooksy & Caracelli, 2005; Mackay, 2002; Trevisan, 2002).

Integration With Organizational Decision-Making

The integration of evaluation with organizational decision-making processes is a key characteristic of organizations that have developed their evaluation capacity and speaks to their capacity to use evaluation judiciously. Typical (low) levels of evaluation capacity for this dimension can be observed in organizations that are fragmented in terms of the location of the evaluation function and the limited communication that takes place between evaluators and program managers. At this level, organizations do not practice needs assessment in the program planning process, and utilization of evaluation results is limited, either because they are unavailable or not considered useful by senior management.

Organizations exhibiting exemplary levels of evaluation capacity for this dimension integrate evaluation in their decision-making processes by inviting evaluators to senior-level administration meetings (Compton et al., 2002), maintaining formal communication channels between evaluators and program managers, incorporating such activities as needs assessment and evaluability assessments to program planning (Brazil, 1999), and positioning the evaluation function where it has the greatest likelihood of contributing to program management decisions (Kegeles, 2005; Owen & Lambert, 1998). The use of evaluation results in budgetary and other types of decisions is also a regular occurrence within the organization.

Evaluation Literacy of Organizational Members

The evaluation literacy of organizational members, or the extent to which organizational members are familiar with evaluation principles and practices, is key to developing the organization's capacity to use evaluation (Milstein et al., 2002). Organizations with typical capacity in this dimension do not have a high degree of evaluation literacy; rather, staff are unfamiliar with evaluation, its principles and potential organizational benefits, and do not usually participate in evaluation-related activities such as sitting on evaluation steering committees due to a lack of time or interest. In addition to this, such participation is not considered for performance appraisal, and evaluation projects are never reviewed or presented to other organizational members in an effort to model good practices.

Organizations that have an exemplary level of evaluation capacity embrace evaluation principles and practices throughout their activities. Organizational members are not only familiar with these principles and practices, but they also participate regularly in evaluation projects and are encouraged to do so by their superiors (Kegeles, 2005; Love, 1998; Newcomer, 2004; Robinson & Cousins, 2004; Sutherland, 2004). Participation in evaluation is considered a key

task of program managers and staff, and therefore they are provided with the time to participate and their performance is appraised partly in terms of their participation in evaluation activities (Newcomer, 2004). Evaluators assess their own work regularly using the same principles they apply to program evaluation in order to model evaluation for their program colleagues and present their findings to a broad organizational audience (Compton et al., 2002).

Interview Process and Findings

The second part of the first phase of the study involved four interviews held with key informants knowledgeable about government evaluation. These individuals were selected based on their current or former positions, either within a government central agency or as consultants working for a number of government clients. The four participants agreed to a one-hour interview focusing on their own definition of evaluation capacity as well as their view of which departments and agencies had high evaluation capacity, and which were building their evaluation capacity (see Appendix C for the Phase 1 interview guide). Further details about the participants and the procedures used are provided in Chapter 3 of the thesis.

A content analysis was performed on the data obtained through the interviews and the results were organized according to the various dimensions of capacity identified. The consolidated data were then used to refine the existing framework according to the process outlined by Leithwood and Montgomery (1987). This process involves identifying the dimensions that describe the innovation being implemented (evaluation capacity in this case), developing the highest stage of the profile, developing the lowest stage of the profile, and completing a description of intermediate stages ensuring manageable steps in the main issue between the highest and lowest stages. In order to add the required detail to the framework, a different format than that used in the preliminary document was adopted. This new format was

borrowed from Cousins et al. (2004) and features one dimension per page presented in matrix form, with sub-dimensions making up the columns and the levels of capacity making up the rows. Three levels of capacity are used here: exemplary capacity, intermediate capacity, and typical or developing capacity.

The findings presented in the revised framework point to a different view of evaluation capacity than had been found through the literature review. A heavy emphasis was placed on organizations' capacity to do evaluation, with the recruitment and technical training of qualified evaluation staff mentioned most often by the participants. This may be a consequence of the current climate for evaluation in the federal government. This issue will be discussed at greater length in Chapter 7, which provides more detailed comments on the findings of this study. The main differences found between the preliminary framework based on the literature and the revised framework that integrates the interview data are discussed in the following paragraphs.

Differences Between Literature and Interview Data

A number of elements included in the preliminary framework were kept in the revised version of the framework because they were found to be important both in the literature review and the interviews conducted in Phase 1. Some of these elements, however, were moved from one section of the framework to another or were reworded to better reflect the structure and language used by governmental organizations in their discussions about evaluation. The five dimensions identified in the preliminary framework will be revisited here along with an outline of the changes made to each of them in the revised framework.

General Comments on Evaluation Capacity

Throughout the interviews, participant shared their thoughts on evaluation capacity with the student researcher, both in terms of what this refers to in organizations, but also to what

extent they felt it was needed in government. A number of participants agreed that evaluation capacity has many facets and that it may be possible to consider it under a number of different points of view. A key comment made by the participants, however, focused on the need for action in organizations wishing to improve their evaluation capacity and the way in which evaluation is used: “building evaluation capacity requires activity at all levels” (Expert 4). One participant qualified this further by stating that capacity building is a process with no concrete ending point. The consequences or results of building an organization’s evaluation capacity were also identified by some participants. For example, one participant stated that:

the first part is helping managers manage and individuals in the organization learn about their own business; the broader, the latter one, the higher-order one, I think is helping organizations manage and plan and look ahead (Expert 2)

Resources for Evaluation

This dimension was divided into two separate dimensions in the revised framework in order to emphasize the importance of both human and organizational resources in an organization’s capacity to do evaluation. The new “Human Resources” dimension (dimension 1) focuses on the evaluation unit and its personnel. The sub-dimensions included in this category include Staffing, Evaluation Logic and Technical Skills, Communications and Interpersonal Skills, Professional Development, and Leadership. Of particular interest are the different types of skills required by evaluators. In addition to the technical expertise often mentioned in the literature, the participants focused heavily on the communication and relationship-building skills needed by evaluators in order to foster the use of evaluation findings within their organizations. In particular, evaluators need to be able to ask difficult questions and to pore through information

as well as act as facilitators throughout the evaluation process. One participant explained the need for these types of skills in organizations:

Even if you have the best technical skills but you can't convince the organization to value what you're doing or co-opt them to participate, then it doesn't matter how good you are (Expert 3).

Furthermore, participants felt that evaluators must perceive evaluation as more than applied social inquiry; they must see the programs, policies or initiatives that they are evaluating as a series of results. They must also convince managers to use the same lens in their work, whether they are involved in evaluation or participating in related activities, such as developing their organization's "Program Activity Architecture"² or writing the "Departmental Performance Report"³. Evaluators need interpersonal skills in order to share their vision of a results-based organization and to foster closer relationships with program staff while avoiding the risk of alienating user groups by putting up barriers of technical jargon.

The issue of leadership is also important to consider within the scope of evaluation human resources. In order to be effective and influential within an organization, an evaluation unit must be headed by a skilled, experienced evaluator who acts as champion and is able to mobilize resources to complete the needed work.

² The Program Activity Architecture (PAA) is an inventory of all the programs and activities undertaken by a department or agency. The programs and activities are depicted in their logical relationship to each other and to the Strategic Outcome(s) to which they contribute (Treasury Board of Canada, Management, Resources, and Results Structure Policy, 2005).

³ Departmental Performance Reports (DPRs), tabled in the fall of each year by the President of Treasury Board on behalf of all federal departments and agencies, present a report on results and accomplishments in order to provide Parliamentarians with knowledge and understanding of the government's stewardship of public resources (Treasury Board of Canada, Results-based Management Lexicon, accessed on July 3rd, 2007 at: http://www.tbs-sct.gc.ca/rma/lex-lex_e.asp)

The new “Organizational Resources” dimension (dimension 2) focuses on some of the points previously included under “Resources for Evaluation”, such as the presence of an adequate evaluation budget. It also incorporates content previously included under the “Evaluation Literacy of Organizational Members” dimension, which was removed from the revised framework. Although evaluation literacy is key to organizational use, its importance in government evaluation is more limited than in other types of organizations and is therefore not considered a key dimension of evaluation capacity for the purposes of this study. However, as the familiarity of organizational staff with evaluation principles can be key to their participation in evaluation studies, a sub-dimension under Organizational Resources was deemed to be appropriate and was included in the revised framework.

A third sub-dimension was created to represent the need for a solid performance measurement system as part of organizational resources for evaluation. The “Ongoing Data Collection” sub-dimension focuses on performance measurement data and how these feed into evaluation studies.

Evaluation Planning and Activities (dimension 3)

This dimension remained in the revised framework, although some of its elements were modified or moved to other dimensions. In particular, participants discussed the need for in-house evaluation whenever possible at great length, stating that in-house evaluation is an important marker of an organization’s capacity to do evaluation. This is reflected through the new “Use of Consultants” sub-dimension. Participants also identified a need for formal or informal discussion on the evaluation process and lessons learned as key to increasing an organization’s evaluation capacity; this was included in the “Information Sharing” sub-dimension. Finally, participants raised the issue of external supports to evaluation, such as

professional associations and quality standards. An organization with high evaluation capacity has access to these external supports and uses them as part of its ongoing evaluation activities. This was also included as a sub-dimension in the framework.

Integration with Organizational Decision-Making (dimension 5)

This dimension was identified as one of the most important factors in evaluation capacity in general, and the most important in an organization's capacity to use evaluation. All interview participants stated that evaluators must be involved in management processes such as the preparation of Memoranda to Cabinet and Treasury Board submissions, since these documents include information on expected results and evaluation issues. As one participant explained,

[evaluators should be connected] to these processes that should be drawing upon evaluation and that are making commitments for evaluation; and if evaluation is plugged into that process chain, that would be a good marker that you've got a department that understands evaluation and values the input of evaluation. (Expert 3)

Another participant described the importance of an organization's capacity to use evaluation by describing its key focus on providing information to be used as part of the decision-making process:

That has to be one of the key tests of capacity, is it part of a chain to lead to action? And if it doesn't lead to action, then, you know, what's the point? I think we need as a group, as a community to really keep that in mind, I do worry sometimes that we focus on capacity for its own sake, and not the fact that evaluation is meant as a decision tool. (Expert 1)

Part of being involved in management processes includes having ready access to the organization's senior managers. Participants emphasized this point -- noted in the

“Organizational Linkages” sub-dimension -- by stating that access to Assistant Deputy Ministers and Deputy Ministers is critical to an organization’s capacity to use evaluation, since senior managers are considered the key clients of the evaluation unit. Ready access to these individuals means “a greater chance of doing evaluation that will make a difference and that will happen at the right time” (Expert 3). According to one participant, senior managers must be involved throughout the evaluation process,

to help ensure that what gets evaluated is what is needed; there’s a dialogue on the front end about issues and then dialogue on the back end in terms of or around the findings, conclusions and recommendations that are tabled and discussed and then finally there’s follow-through on that (Expert 2).

In other words, if evaluation is meant to influence decision-making, the decision makers themselves must take an interest in what is being evaluated and have the intention of using that information in the decision-making process: “what happens to the information? We’re all about influencing decision-making, and if it doesn’t do that, what does that mean?” (Expert 1).

The result of this interest is presented in the “Decision Support” sub-dimension, which describes how evaluation findings can be used by senior management and also shows awareness of evaluation services across the organization through demand for these services.

Finally, the location of the evaluation function and the performance of needs assessment, two elements included under this dimension in the preliminary framework, were modified in the revised version in order to better reflect the unique context of Canadian federal government organizations. The location of the evaluation function was thought to be included in the sub-dimension dealing with organizational linkages, and the performance of needs assessment was included in the Evaluation Plan sub-dimension.

Other Dimensions

One of the dimensions included in the preliminary framework, “Assessment of ECB Efforts”, was not mentioned by the interview participants. It is possible that this dimension may not reflect the experience of participants in terms of ECB. Although this dimension was not included in the revised version of the framework for this reason, some of its elements were modified and integrated to other dimensions. For example, the element “formal or informal processes to discuss lessons learned during evaluation” was slightly modified and moved to the Evaluation Planning and Activities dimension.

Two new dimensions emerged through the interview data content analysis. Participants stated that the capacity to use evaluation is also dependent on certain organizational characteristics (dimension 4). A results-management orientation, evidenced through the presence of clear ideas about organizational purposes, goals and mandate, was identified as key to government organizations. This requires openness on the part of managers to improve their programs, as well as an understanding both among managers and staff that they need to consider whether they are “doing the right things vs. doing things right” (Expert 1), the latter too often being a major issue raised in evaluation studies but better left to audits. In a nutshell, the results-management orientation refers to “a knowledge-based type of organization that is prepared to conduct studies and monitor performance of activities, programs and policies, but also prepared to make changes based on information” (Expert 2).

A mature organizational structure is also required to achieve a high level of evaluation capacity. This element was characterized by participants as policies, functioning, leadership, governance structure and good planning, all working towards maximizing the relevance and use of evaluation.

Finally, information sharing was identified by participants as being critical to evaluation use. In addition to the usual vertical channels in place in government organizations, participants expressed a need for increased horizontal information sharing about evaluation.

The learning benefits of evaluation (dimension 6), as evidenced through instrumental/conceptual use and process use were also raised as an important dimension of evaluation capacity. Instrumental and conceptual use feature the use of the evaluation findings to foster changes in program design or changes in stakeholders' attitudes about the program evaluated. Process use focuses on the behavioural and cognitive changes that occur through stakeholders' proximity to evaluation.

Summary of Revisions for Phase 1

The revised framework, like the preliminary framework, divides the dimensions of evaluation capacity into two sections, capacity to do evaluation and capacity to use evaluation. The revised framework is displayed differently than the preliminary document in order to accommodate a greater level of detail. Although most of the dimensions are relatively similar from one version to the next, several changes were made based on interview results. The six dimensions included in the framework are now called Human Resources, Organizational Resources, Evaluation Planning and Activities, Organizational Characteristics, Integration with Organizational Decision-Making and Learning Benefits. Two dimensions included in the preliminary framework were not mentioned by the participants: Assessment of ECB Efforts and Evaluation Literacy of Organizational Members are not included in the revised version as individual dimensions, although the elements included under these were moved to other dimensions of the framework to retain their essence. Table 2 below includes the revised framework and summarizes the findings of Phase 1.

Table 3: Phase 1 Framework

CAPACITY TO DO EVALUATION

1. Human Resources					
Level	Staffing	Evaluation Logic and Technical Skills	Communications and Interpersonal Skills	Professional Development	Leadership
Exemplary Capacity	<ul style="list-style-type: none"> Evaluation unit is fully staffed Appropriate balance of senior and junior evaluators 	<ul style="list-style-type: none"> All staff members possess the necessary evaluation knowledge and skill to carry out the unit's projects 	<ul style="list-style-type: none"> All staff members possess the appropriate communication and relationship-building skills and experience to carry out their projects 	<ul style="list-style-type: none"> All staff members engage in professional development activities directly related to their work at least once per year Learning plans are developed for all staff members annually 	<ul style="list-style-type: none"> Evaluation unit headed by individual with strong evaluation expertise Leader is known to be champion of evaluation across organization Staff members are comfortable with taking direction from leader
Intermediate Capacity	<ul style="list-style-type: none"> Evaluation unit is partially staffed Appropriate balance of senior and junior evaluators 	<ul style="list-style-type: none"> Most staff members have the necessary evaluation knowledge and skills to carry out the unit's projects 	<ul style="list-style-type: none"> Staff members are gaining experience and receiving training to develop communication and relationship-building skills 	<ul style="list-style-type: none"> Most staff members engage in professional development activities directly related to their work at least once a year Learning plans are developed for all staff members annually 	<ul style="list-style-type: none"> Unit headed by individual with general evaluation background Leader promotes evaluation within organization Staff members are mostly comfortable with taking direction from leader
Typical or Developing Capacity	<ul style="list-style-type: none"> Vacant positions at all levels within the evaluation unit 	<ul style="list-style-type: none"> Staff members do not have a background in technical aspects of evaluation (e.g., logic modeling) Significant demand for prof. development 	<ul style="list-style-type: none"> Staff members generally do not engage in relationship-building activities with stakeholders Communication with stakeholders is limited 	<ul style="list-style-type: none"> Some staff members occasionally engage in professional development activities Some discussion of learning goals takes place 	<ul style="list-style-type: none"> Unit headed by individual with some knowledge and experience in evaluation

CAPACITY TO DO EVALUATION

2. Organizational Resources			
Level	Budget	Ongoing Data Collection	Evaluation Literacy
Exemplary Capacity	<ul style="list-style-type: none"> Evaluation budget ensured through A-base funds and earmarked for evaluation projects (i.e., cannot be diverted to other corporate priorities) Evaluation budget sufficient to complete activities outlined in multi-year evaluation plan 	<ul style="list-style-type: none"> Performance measurement system in place across organization Performance data feeds directly into evaluation studies 	<ul style="list-style-type: none"> Organizational members at all levels understand the principles and practices of evaluation Organizational members are involved throughout the evaluation process (e.g., identifying evaluation issues, developing data collection instruments, interpreting results)
Intermediate Capacity	<ul style="list-style-type: none"> Evaluation budget ensured through a combination of stable funding sources (i.e., program funding and corporate evaluation funding) Evaluation budget adequate to complete activities outlined in multi-year evaluation plan 	<ul style="list-style-type: none"> Performance measurement system in place for major programs or activities Performance data can be adapted to suit the informational requirements of evaluation studies 	<ul style="list-style-type: none"> Organizational members are familiar with the principles behind evaluation Organizational members participate in some evaluation activities (e.g., reviewing data collection instruments, development of evaluation framework)
Typical or Developing Capacity	<ul style="list-style-type: none"> Evaluation budget provided on a case-by-case basis for each new project Evaluation budget sufficient to complete only activities planned for current year; future years are uncertain 	<ul style="list-style-type: none"> Performance measurement is done on a program-by-program basis Performance data is difficult to integrate into evaluation studies 	<ul style="list-style-type: none"> Little or no involvement from program staff and managers (e.g., some input during evaluation framework development process, or review of evaluation reports)

CAPACITY TO DO EVALUATION

3. Evaluation Planning and Activities				
Level	Evaluation Plan	Use of Consultants	Information Sharing	External Supports
Exemplary Capacity	<ul style="list-style-type: none"> • Existence of organization-wide, multi-year evaluation plan • Evaluation plan developed in consultation with all senior managers and includes needs assessment exercise • Evaluation plan approved by senior management committee • Evaluation plan includes thorough risk assessment process 	<ul style="list-style-type: none"> • Appropriate balance of evaluations designed and conducted by evaluation staff, or by consultants for specific expertise (e.g., when a project's scope is too large for the evaluation unit's resources) 	<ul style="list-style-type: none"> • Major decisions on evaluation projects are made collaboratively to benefit from all team members' knowledge and experience • Formal or informal processes to discuss lessons learned during evaluations are in place and involve the entire organization (e.g., seminars, brown-bag lunch sessions, brochures on recent studies by theme) 	<ul style="list-style-type: none"> • Evaluators make frequent use of external supports to evaluation such as professional association, published quality standards, etc.
Intermediate Capacity	<ul style="list-style-type: none"> • Evaluation five-year plan (at least) exists • Evaluation plan developed in consultation with most senior managers • Evaluation plan includes some assessment of risk 	<ul style="list-style-type: none"> • Medium and large evaluations are designed in-house and conducted by external consultants • Evaluation staff involved in conducting the evaluation (i.e., contribute to some of the field work done by consultants) • Smaller studies are conducted in-house by evaluation staff 	<ul style="list-style-type: none"> • Evaluators share their progress with their colleagues at regular team meetings • Lessons learned through evaluations are shared with organizational members directly involved with the program (e.g., memos, formal presentation of report) 	<ul style="list-style-type: none"> • Evaluators make some use of external supports to evaluation such as professional association, published quality standards, etc.
Typical or Developing Capacity	<ul style="list-style-type: none"> • Evaluation plan exists for one or two years • Evaluation plan developed in consultation with some program managers or senior managers 	<ul style="list-style-type: none"> • Virtually all evaluations are designed and conducted by external consultants • Evaluation staff is mainly involved in managing contracts and overseeing the work done 	<ul style="list-style-type: none"> • Evaluators share their progress with their supervisors and other team members in a sporadic, informal manner • Evaluation projects are not discussed at the organizational level once completed; evaluation reports disseminated only to internal evaluation committee 	<ul style="list-style-type: none"> • Evaluators have access to basic external supports, such as a professional association or published quality standards but do not often make use of them

CAPACITY TO USE EVALUATION

4. Organizational Characteristics			
Level	Results-Management Orientation	Organizational Infrastructure	Information Sharing
Exemplary Capacity	<ul style="list-style-type: none"> Senior managers promote a results-management orientation for the entire organization Organizational members share clear ideas about organizational purpose and goals All programs have a clear results chain (i.e., logic model) Program staff are directly responsible for the development of RMAFs and other performance management activities 	<ul style="list-style-type: none"> Mature organizational infrastructure demonstrated through the implementation of organizational policies on evaluation and performance measurement, a stable governance structure, a structured planning process that includes consideration of evaluation issues and findings 	<ul style="list-style-type: none"> Information sharing occurs vertically and horizontally throughout organization
Intermediate Capacity	<ul style="list-style-type: none"> Organizational outcomes or expected results are only outlined in official documentation but are not included in communications from senior managers Some programs are engaged in logic modeling or have completed the logic modeling process Program staff contribute to the development of RMAFs and other performance management activities 	<ul style="list-style-type: none"> Organizational infrastructure showing some maturity (e.g., clear governance structure and stability in senior management ranks; organizational policy on evaluation has been developed and implemented) 	<ul style="list-style-type: none"> Information is shared only vertically and not between program areas
Typical or Developing Capacity	<ul style="list-style-type: none"> Organizational outcomes or expected results are not articulated clearly for all organizational members Some programs are engaged in logic modeling Program staff not involved in the development of RMAFs and other performance management activities 	<ul style="list-style-type: none"> Organizational infrastructure still under development (e.g., no clear governance structure, no policies on evaluation are in place, no structured planning process) 	<ul style="list-style-type: none"> Little information is shared between or across levels of the organization

CAPACITY TO USE EVALUATION

5. Integration with Organizational Decision-Making			
Level	Management Processes	Organizational Linkages	Decision Support
Exemplary Capacity	<ul style="list-style-type: none"> Program and policy staff always request the involvement of evaluators in management processes such as MC and TB submission preparation 	<ul style="list-style-type: none"> Evaluators in constant communication with program managers through formal or informal ties Ready access to ADM and DM on all aspects of evaluation; clear interest in evaluation information demonstrated by ADM and DM 	<ul style="list-style-type: none"> Evaluation findings considered in budget allocation and other high-level organizational and policy decisions Demand for evaluation services originates from all levels of the organization
Intermediate Capacity	<ul style="list-style-type: none"> Program and policy staff consult evaluators for certain management processes (e.g., when results information or specific feedback concerning evaluation is required) 	<ul style="list-style-type: none"> Evaluators communicate regularly with their program clients DM and ADM receive regular reports about evaluation activities but are not directly involved in evaluation 	<ul style="list-style-type: none"> Evaluation findings usually considered in program management decisions and some policy decisions Program managers are interested in and use evaluation as a management support tool (i.e., evaluation as provider of ongoing management information)
Typical or Developing Capacity	<ul style="list-style-type: none"> Evaluation unit operates separately from program units and is not generally involved in management processes; program and policy staff unaware of the potential contributions of evaluation staff 	<ul style="list-style-type: none"> Evaluators communicate with program clients on specific issues related to projects Senior managers are made aware of evaluation findings only through one or two official mechanisms (e.g., internal evaluation committee) 	<ul style="list-style-type: none"> Little consideration of evaluation findings in organizational and policy decisions No specific demand for evaluation services other than to meet TB requirements

CAPACITY TO USE EVALUATION

6. Learning Benefits	
Level	Instrumental/Conceptual Use
Exemplary Capacity	<ul style="list-style-type: none"> • Evaluation findings are used consistently as a basis for action and change (i.e., evaluation recommendations are appropriate and implemented in a timely manner) • Evaluation findings and reports often have an impact on stakeholders' understanding and attitudes about programs
Intermediate Capacity	<ul style="list-style-type: none"> • Evaluation findings are sometimes used as a basis for action and change (i.e., evaluation recommendations are sometimes implemented) • Evaluation findings and reports can have an impact on stakeholders' understanding and attitudes about programs
Typical or Developing Capacity	<ul style="list-style-type: none"> • Evaluation findings are rarely used as a basis for action and change (i.e., evaluation recommendations are usually not implemented) • Evaluation findings are reports rarely have an impact on stakeholders' understanding and attitudes about programs
	Process Use
	<ul style="list-style-type: none"> • Strong evidence of behavioural or cognitive changes occurring in stakeholders by virtue of their proximity to evaluation (e.g., stakeholders apply evaluation logic to other organizational issues)
	<ul style="list-style-type: none"> • Some evidence of behavioural or cognitive changes occurring in stakeholders by virtue of their proximity to evaluation (e.g., stakeholders apply evaluation logic to other organizational issues)
	<ul style="list-style-type: none"> • No evidence of behavioural or cognitive changes occurring in stakeholders by virtue of their proximity to evaluation (e.g., stakeholders apply evaluation logic to other organizational issues)

It can be noted in the revised framework that the number of levels varies from one sub-dimension to the next. This is due mainly to the fact that some sub-dimensions have more or less complex elements and that it may take more or less effort to move between some levels compared to others. For example, three of the five Human Resources sub-dimensions have been divided into four levels in order to properly capture all of the changes that need to be implemented in order to move from typical or developing capacity to exemplary capacity.

Because each of the sub-dimensions includes a number of different descriptive elements, it may be possible for organizations to achieve higher capacity on some of these elements and not on some of the others. In this case, establishing the actual level of capacity of an organization on any of the six dimensions would be difficult. It is important to remember that the revised framework is meant to provide guidance to organizations interested in increasing their evaluation capacity by suggesting avenues for improvement based on their own assessment. Seen in this light, it is possible for organizations to recognize that they are doing particularly well on one element (depicted by a bullet point in the framework) within a level but need to improve on other elements before achieving a specific level of capacity.

Markers of Organizational Evaluation Capacity

In order to better identify the various dimensions of evaluation capacity and their elements, the interview participants were asked what some of the key markers of evaluation capacity might be (i.e., “What would you look for in an organization in order to assess its evaluation capacity?”). These markers were integrated to a large extent in the revised framework of evaluation capacity. However, it may be interesting to outline these markers here, as they constitute useful indicators of evaluation capacity and may be of use in a diagnostic instrument

to be used by organizations (although the development of such an instrument is beyond the scope of the present study).

An organization's capacity to use evaluation can be inferred from its management processes and supporting documentation. For example, an organization that produces a Management Accountability Framework (MAF) that shows quality, impact of evaluation and implementation of evaluation recommendations likely has high capacity to use evaluation. In the same sense, an evaluation's influence on policy can be seen in various changes or in the implementation of evaluation findings and once again speaks to the ability of an organization to use evaluation results. As mentioned previously, Memoranda to Cabinet and Treasury Board submissions prepared with the help of evaluation professionals reflect a specific interest in results information and show that evaluation is considered throughout a program's lifecycle.

Other indicators of an organization's evaluation capacity focus on the specific findings of evaluations and their use. Increased effectiveness of a program due to the implementation of evaluation results is a clear indicator of evaluation use. An increase in the demand for evaluation services from managers and staff is also a clear marker of developing evaluation capacity. Unlike the previous set of indicators, which speak to an organization's capacity to use at a macro level, these indicators focus more closely on individual programs and their managers and speak to an organization's ability to learn about its programs and their effectiveness.

Capacity of Specific Departments and Agencies

In an effort to illustrate the key dimensions of evaluation capacity, participants in the first phase of the study were asked to name specific departments or agencies with high, developing, or low evaluation capacity. The findings obtained in the interviews are presented below, although the names of the organizations are withheld in order to protect the individuals

associated with these organizations. Some of these organizations were subsequently selected for the third phase of the study, described in Chapter 6.

High Capacity Organizations

Two organizations were identified by all of the interview respondents as having particularly high evaluation capacity. The first organization was hailed as an example to other organizations, having made deliberate efforts to build evaluation capacity both in the evaluation unit and across the department. They are known for their practical approach, linking evaluation with results-based management. However, the participants were unsure about the organization's capacity to use evaluation, "whether hard-hitting stuff stays as hard". The second organization was identified by participants because of its reputation as a producer of high-quality evaluation studies and has been recognized for the rigour of its evaluation work and for its investment in technical capacity. However, as for the first organization nominated, participants wondered about this organization's capacity to use evaluation, especially given the time required to conduct such high-caliber evaluation studies. It appears therefore that although these two organizations have high capacity to do evaluation, they may still be building their capacity to use evaluation.

Intermediate Capacity Organizations

Some of the organizations identified by the interview participants had high evaluation capacity on some elements but lower on others. In general, these organizations have worked towards building strong evaluation units made up of experienced professionals charged with establishing linkages with senior management. For example, the evaluators in one organization have focused heavily over the past few years on responding to the requirements of senior managers and on working closely with the Treasury Board Secretariat. Another one of these organizations has a fairly mature evaluation function that is known to be well-connected to

senior management, especially in terms of preparing Memoranda to Cabinet and Treasury Board submissions. A third organization has been increasing its internal capacity and establishing linkages in the same way, and the evaluators in a fourth organization have increasingly been conducting in-house work and have insinuated themselves into management processes.

Typical or Developing Capacity Organizations

Because so many government organizations have fairly low evaluation capacity, there was little agreement among participants on the actual organizations identified in the interviews. A few participants mentioned that many small agencies do not have an evaluation function and so constitute the government organizations with the lowest evaluation capacity. Among the larger organizations, two were identified as being in this group. Although one of these has fairly good capacity to do evaluation, the environment in which it operates makes it difficult to develop its capacity to use evaluation. The politically-charged context in which this organization operates, among others, makes evaluation only a small component of complex policy decisions. Along the same lines, there has been little evidence of use at another one of the organizations identified, given that the rationale of most of its programs is difficult to question. Other problems relating to evaluation capacity are linked to building the evaluation function itself and so fall within the realm of capacity to do evaluation. Two other organizations are still building their evaluation teams, which are made up mostly of junior officers. It is therefore difficult to make those important connections to senior managers and to become more involved in their processes, given that the evaluators themselves need more experience and more training in order to produce quality evaluations that are comparable to those produced by more mature evaluation shops.

Selection of Organizations for Phase 3 of the Study

Five organizations were selected to participate in Phase 3 of the study, based on the interview data collected in Phase 1. The two organizations with high evaluation capacity were selected because of the ability of their managers and staff to comment on their evaluation capacity building efforts and because it was anticipated that these organizations are at least somewhat experienced in using evaluation. One of the organizations thought to have intermediate evaluation capacity was selected to participate in the last phase of the study because of its recent efforts at building capacity and because of the development of a new identity for the organization. It was thought that evaluators and program managers alike would be able to comment on the evaluation capacity framework based on these recent experiences. Finally, two typical or developing capacity organizations were selected to participate in the study. One of these has only recently begun building its capacity to do evaluation after a significant change in management. Another interesting aspect of this organization is that it has just started moving from an operational to a results-based management perspective and so the efforts of staff in applying this new lens to their evaluation process may prove to be interesting in terms of validating the framework's dimensions. The second typical or developing capacity organization was selected based on the fact that it is also trying to build its evaluation function from the ground up and experiencing some difficulties in attracting and retaining qualified technical staff. All but the last organization selected participated in Phase 3, as will be described in Chapter 6. One of the typical or developing capacity organizations declined an invitation to participate because of internal issues related to workload and staff turnover.

Chapter 5 – Findings From Phase 2

This chapter summarizes the findings obtained in the second phase of the study. This phase focused on refining the framework developed in the first phase and featured interviews with the same four participants who provided the data used in part to develop the framework. In the second round of interviews, participants were asked to look at the framework and to comment on its format, content, and potential use in a Canadian federal government organization. Appendix D presents the interview protocol used in this phase of the study.

The following paragraphs identify the changes made to the framework based on the second round interview data. The changes made to the format, structure and overall appearance of the framework will be presented in the first section of the chapter. The next section will focus more closely on the individual dimensions and sub-dimensions and the changes that were made in each of these. A summary of the framework is presented in Figure 2, and a finalized version of the framework is presented in Table 3 below.

Figure 3: Summary of Phase 2 Framework

EVALUATION CAPACITY IN CANADIAN FEDERAL GOVERNMENT ORGANIZATIONS
FRAMEWORK SUMMARY

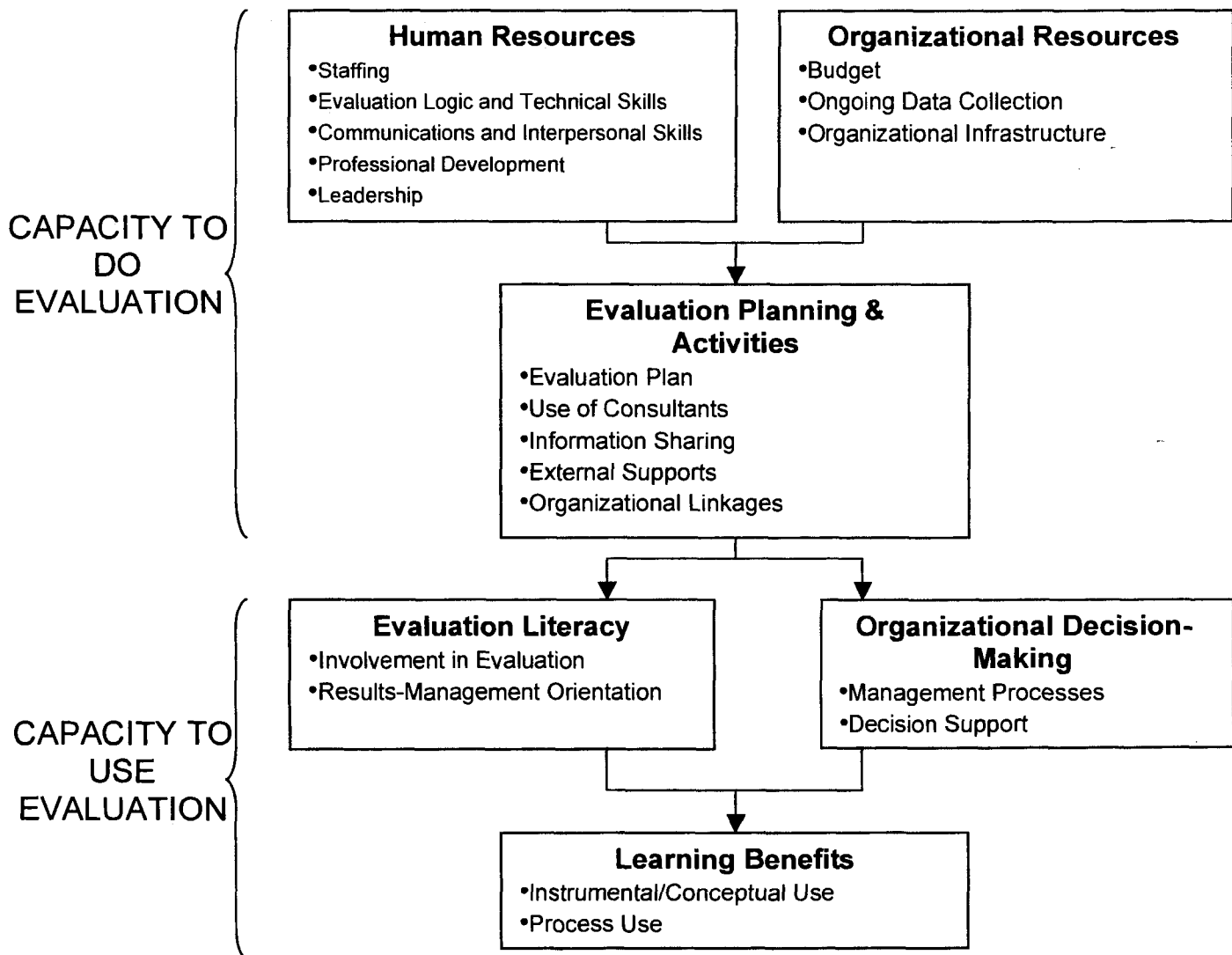


Table 4: Phase 2 Framework

CAPACITY TO DO EVALUATION

		1. Human Resources				
		1.1 Staffing	1.2 Evaluation Logic and Technical Skills	1.3 Communications and Interpersonal Skills	1.4 Professional Development	1.5 Leadership
Level	Exemplary Capacity	<ul style="list-style-type: none"> Evaluation unit is fully staffed Appropriate balance of senior and junior evaluators, given organizational resources and requirements Career progression process is in place to facilitate promotions within unit (through developmental opportunities) 	<ul style="list-style-type: none"> Evaluators are able to recognize strategic uses of evaluation methods in supporting organizational decision-making Evaluators are able to identify key evaluation issues and operationalize these through the selection of appropriate evaluation methods Evaluators are able to extract key findings from evaluations and draw linkages to organizational priorities Evaluators are able to detect obstacles to project successes and develop solutions, and analyze and reconcile competing priorities 	<ul style="list-style-type: none"> Evaluators effectively manage interpersonal relationships within and across projects; promote commitment to maintaining and building client trust and are able to diffuse potentially confrontational situations Evaluators promote and deliver open, clear and transparent messages across the organization and on its behalf 	<ul style="list-style-type: none"> All staff members engage in external professional development activities directly related to their work at least once per year (e.g., conferences) Organization develops its own internal professional development activities (e.g., brown bag sessions, peer review processes, etc) 	<ul style="list-style-type: none"> Evaluation unit headed by individual with strong evaluation and management background Leader effectively translates senior managers' needs into concrete project plans Leader guides, mentors or coaches team members Leader establishes and ensures that goals related to quality, productivity and timelines are achieved

<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> • Evaluation unit is partially staffed • Imbalance of senior and junior evaluators • Career progression process is available for some levels of evaluation professionals wishing to gain experience in their field 	<ul style="list-style-type: none"> • Evaluators are able to identify key issues and apply sound methodologies to evaluate programs, analyze and interpret findings and propose feasible and useful recommendations • Evaluators have essential project management competencies 	<ul style="list-style-type: none"> • Evaluators manage projects in a manner that fosters strong relationships and client trust (e.g., deliver sensitive messages with tact) • Evaluators are able to adapt the style and contents of their communications to the needs of the audience 	<ul style="list-style-type: none"> • Most staff members engage in external professional development activities directly related to their work once every few years • Some internal learning resources are made available to staff members (e.g., journals, books) 	<ul style="list-style-type: none"> • Unit headed by individual with some evaluation and management experience • Leader explains senior management needs to lead evaluators • Leader delegates work appropriately and maximizes team productivity
<p>Low or Developing Capacity</p>	<ul style="list-style-type: none"> • Some positions vacant at all levels within the evaluation unit • Career progression process does not exist for evaluators other than the traditional competitions when positions become vacant 	<ul style="list-style-type: none"> • Evaluators possess basic skills required to identify and apply appropriate data gathering and analysis techniques, extract critical pieces of information from data and draw sound conclusions based on evidence 	<ul style="list-style-type: none"> • Evaluators work towards developing good relationships and client trust • Evaluators are able to present information logically within specific contexts (i.e., evaluation steering committee meetings) 	<ul style="list-style-type: none"> • Some staff members engage in professional development activities delivered by external organizations on occasion 	<ul style="list-style-type: none"> • Unit headed by individual who is new to the area of evaluation and/or has limited management experience • Leader does not participate in senior management discussions • Leader coordinates team activities but is not involved in ensuring quality or productivity

CAPACITY TO DO EVALUATION

Level	2. Organizational Resources		
	2.1 Budget	2.2 Ongoing Data Collection	2.3 Organizational Infrastructure
Exemplary Capacity	<ul style="list-style-type: none"> Evaluation budget ensured through continuing funding to organization and earmarked for evaluation projects (i.e., cannot be diverted to other corporate priorities) 	<ul style="list-style-type: none"> Performance measurement system fully implemented across organization High-quality information is collected by the PM system Performance data feeds directly into results-based management (including evaluation studies) 	<ul style="list-style-type: none"> The organization has a stable governance structure and clear accountability lines for results Organizational policies on evaluation and performance measurement have been developed and implemented Mechanisms are in place to ensure that performance data feed into a structured planning and reporting process (e.g., use of evaluation in RPP, budgeting, etc) Existence of organizational supports that provide needed services to evaluators in a competent and timely manner, allowing them in turn to produce timely and useful evaluation reports (i.e., procurement, communications, HR, etc)
Intermediate Capacity	<ul style="list-style-type: none"> Evaluation budget ensured through a combination of stable funding sources (i.e., funding obtained through TB submissions and corporate evaluation funding) 	<ul style="list-style-type: none"> Performance measurement system in place for major programs or activities Incomplete implementation of PM system (i.e., sporadic data collection, missing variables, etc) Performance data can be adapted to suit the informational requirements of results-based management (including evaluation studies) 	<ul style="list-style-type: none"> Organizational infrastructure showing some maturity (e.g., clear governance structure and stability in senior management ranks; organizational policy on evaluation has been developed and implemented) Existence of a structured planning process that includes consideration of evaluation issues and findings (e.g., use of evaluation in RPP, budgeting, etc) Organizational supports do not always provide needed services to evaluators in a timely manner and results in delays for evaluation projects
Low or Developing Capacity	<ul style="list-style-type: none"> Evaluation budget provided on a case-by-case basis for each new project 	<ul style="list-style-type: none"> Performance measurement is done on a program-by-program basis Ad hoc implementation of performance measures Performance data is difficult to integrate into results-based management (including evaluation studies) 	<ul style="list-style-type: none"> Organizational infrastructure still under development (e.g., no clear governance structure, no policies on evaluation are in place, no structured planning process) Organizational supports often fail to provide needed services to evaluators

CAPACITY TO DO EVALUATION

Level	3. Evaluation Planning and Activities				3.5 Organizational Linkages
Exemplary Capacity	3.1 Evaluation Plan	3.2 Use of Consultants	3.3 Information Sharing	3.4 External Supports	3.5 Organizational Linkages
	<ul style="list-style-type: none"> • Existence of organization-wide, multi-year evaluation plan • Evaluation plan developed in consultation with all senior managers and includes needs assessment exercise • Evaluation plan includes thorough risk assessment process • Ongoing intelligence gathering allows for changes to be made to plan when necessary • Evaluation plan includes systematic review of evaluation unit itself (i.e., establishing and measuring service standards for the unit) 	<ul style="list-style-type: none"> • Appropriate balance of evaluations designed and conducted by evaluation staff, or by consultants for specific expertise (e.g., when a project's scope is too large for the evaluation unit's resources or when dealing with interdepartmental evaluations) 	<ul style="list-style-type: none"> • Major decisions on evaluation projects are made collaboratively within the unit to benefit from all team members' knowledge and experience • Evaluators actively gather information on new developments in policy and strategic planning 	<ul style="list-style-type: none"> • Evaluators make frequent use of external supports to evaluation such as professional association, published quality standards, etc. • Evaluators are actively involved in broadening their external networks by liaising with evaluators in other organizations, engaging academic experts, and gathering information on activities of OAG, TBS, Privy Council, etc • External stakeholders responsible for the delivery of federal government programs are involved in evaluation activities 	<ul style="list-style-type: none"> • Evaluators in regular contact with program managers through formal or informal ties • Ready access to ADM and DM on all aspects of evaluation; clear interest in evaluation information demonstrated by ADM and DM • Evaluation unit located in close proximity to key organizational areas such as policy development, strategic planning and performance measurement units

<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> • Evaluation five-year plan exists • Evaluation plan developed in consultation with most senior managers • Evaluation plan includes some assessment of risk • Provisions made for reviewing the plan on an ongoing basis 	<ul style="list-style-type: none"> • Medium and large evaluations are designed by in-house and conducted by external consultants • Evaluation staff involved in conducting the evaluation (i.e., contribute to some of the field work done by consultants) • Smaller studies are conducted in-house by evaluation staff 	<ul style="list-style-type: none"> • Evaluators share their progress with their colleagues at regular team meetings • Evaluators are generally aware of new developments in policy and strategic planning 	<ul style="list-style-type: none"> • Evaluators make some use of external supports to evaluation such as professional association, published quality standards, etc. • Evaluators keep themselves informed of new developments in external organizations of interest (i.e., universities, other departments and agencies, central agencies, etc) • External stakeholders involved in the delivery of federal government programs are kept informed of evaluation activities 	<ul style="list-style-type: none"> • Evaluators communicate with their program clients through ongoing, formal mechanisms • DM and ADM receive regular reports about evaluation activities but are not directly involved in evaluation • Evaluation unit located in close proximity to some key organizational areas such as policy development, strategic planning and performance measurement
<p>Low or Developing Capacity</p>	<ul style="list-style-type: none"> • Evaluation plan exists for one or two years • Evaluation plan developed in consultation with some program managers or senior managers • No consideration of risk in planning evaluation schedule 	<ul style="list-style-type: none"> • Virtually all evaluations are designed and conducted by external consultants • Evaluation staff is mainly involved in managing contracts and overseeing the work done 	<ul style="list-style-type: none"> • Evaluators share their progress with their supervisors and other team members in a sporadic, informal manner • Evaluators are not aware of new developments in policy and strategic planning 	<ul style="list-style-type: none"> • Evaluators have access to basic external supports, such as a professional association or published quality standards but do not often make use of them • Evaluators do not generally liaise with external organizations or experts • External stakeholders are not informed of evaluation activities 	<ul style="list-style-type: none"> • Evaluators communicate with program clients on specific issues related to projects • DM and ADM are made aware of evaluation findings only through formal requirements (e.g., internal evaluation committee) • Evaluation unit is removed (either physically or structurally) from key organizational areas such as policy development, strategic planning and performance measurement

CAPACITY TO USE EVALUATION

Level	4.1 Involvement in Evaluation	4.2 Results-Management Orientation
<p>Exemplary Capacity</p>	<p>4. Evaluation Literacy</p> <ul style="list-style-type: none"> Organizational members at all levels understand the purpose of evaluation and it supports the organizational mandate Program managers and other staff members are closely involved at key points in the evaluation process (e.g., review identified issues and provide feedback, facilitate data collection opportunities, review draft evaluation reports, etc) 	<ul style="list-style-type: none"> Senior managers promote a results-management orientation for the entire organization and make it a priority by providing time and resources Organizational members share clear ideas about organizational purpose and goals through formal and informal mechanisms (e.g., strategic planning sessions, retreats, regular meetings, brown bag lunch sessions, etc) All programs have a clear results chain (i.e., logic model) Program managers take the lead for the development of RMAFs and other performance management activities; evaluators assist when needed Organizational outcomes or expected results are only outlined in official documentation but are not included in communications from senior managers
<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> Organizational members are familiar with the general principles of evaluation and how it can help them in their work Program managers sit on Evaluation Steering Committees and provide program-related feedback on report drafts 	<ul style="list-style-type: none"> Organizational members share clear ideas about organizational purpose and goals through formal mechanisms (i.e., strategic planning sessions, meetings, etc) Some programs have a clear results chain (i.e., logic model) Program managers work with evaluators in the development of RMAFs and other performance measurement activities, but evaluators lead these projects
<p>Low or Developing Capacity</p>	<ul style="list-style-type: none"> Little or no awareness of evaluation or its purpose within larger organizational context Little or no involvement from program staff and managers (i.e., brief comments on draft evaluation reports) 	<ul style="list-style-type: none"> Organizational outcomes or expected results are not articulated clearly for all organizational members; most are not aware of results management principles and practices Some programs are engaged in developing results chains such as logic models Program managers not involved in the development of RMAFs and other performance management activities; evaluators conduct these processes with little input from programs

CAPACITY TO USE EVALUATION

5. Integration with Organizational Decision-Making	
Level	5.2 Decision Support
Exemplary Capacity	<p style="text-align: center;">5.1 Management Processes</p> <ul style="list-style-type: none"> • Program and policy staff routinely request the involvement of evaluators in management processes such as MC and TB submission preparation
Intermediate Capacity	<ul style="list-style-type: none"> • Program and policy staff are aware of evaluation unit and ask questions about its services
Low or Developing Capacity	<ul style="list-style-type: none"> • Evaluation findings usually considered in program management decisions and some policy decisions • Program managers are interested in and use evaluation as a management support tool (i.e., evaluation as provider of ongoing management information) • Little consideration of evaluation findings in organizational and policy decisions • No specific demand for evaluation services other than to meet TB requirements

CAPACITY TO USE EVALUATION

6. Learning Benefits	
Level	6.2 Process Use
<p>Exemplary Capacity</p>	<p style="text-align: center;">6.1 Instrumental/Conceptual Use</p> <ul style="list-style-type: none"> • Evaluation findings are used consistently as a basis for action and change (i.e., evaluation recommendations are appropriate and implemented in a timely manner) • Evaluation findings and reports often have an impact on stakeholders' understanding and attitudes about programs
<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> • Evaluation findings are sometimes used as a basis for action and change (i.e., evaluation recommendations are sometimes implemented) • Evaluation findings and reports can have an impact on stakeholders' understanding and attitudes about programs
<p>Low or Developing Capacity</p>	<ul style="list-style-type: none"> • Evaluation findings are rarely used as a basis for action and change (i.e., evaluation recommendations are usually not implemented) • Evaluation findings are reports rarely have an impact on stakeholders' understanding and attitudes about programs

General Changes Made to Framework

The initial comments made by interview participants while reviewing the framework focused heavily on its general appearance and format. More specifically, the participants highlighted a need for a front page summary and commented on the names and number of the levels used in the framework.

Front Page Summary

The participants all felt that the format of the framework itself (i.e., one dimension per page, divided into sub-dimensions) was clear and easy to understand. Some of the participants felt, however, that a front page summarizing the framework would be helpful in getting a general sense of how the dimensions fit together and in providing a glimpse into the framework before getting into the details: “people may focus on the individual boxes and lose the overall perspective” (Expert 1). A summary showing the six dimensions and all of the sub-dimensions was therefore added as front page.

Names of Levels

The participants were somewhat concerned with one of the names used in describing the three levels of evaluation capacity presented for each of the six dimensions. Although “exemplary capacity” and “intermediate capacity” were judged to be suitable in representing each level, the use of the word “typical” worried one participant in particular, who felt that this was not strong enough wording and instead preferred the use of the words “low capacity”:

If this was low (capacity), and if this was my department, then I could use this as rationale or justification for why I need a staff, why I need resources, why I need to move forward, because we’re at low capacity and we need to move up. But if we’re at typical, typical implies okay, that (...) we’re average, we don’t need to worry, and it’s just if you

want to be better, you do this. But it doesn't convey to me that there's a problem (Expert 3).

Other participants also felt that the use of the word "typical" suggested a level that is considered normal, and encouraged more direct wording: "...call a spade a spade" (Expert 4). One other suggested the use of "needs improvement" to convey that there are steps that should be taken to improve the organization's evaluation capacity.

Based on all of these comments, the wording used in the framework was revised to "low or developing capacity". It is expected that the use of these more specific terms will reflect a situation in which an organization has little or no evaluation capacity without some of the more negative connotations that may be attributed to the use of only the words "low capacity".

It should be noted, however, that the descriptions included in "low capacity" level remain fairly typical of many government organizations. In some cases, the elements included in the descriptions may be considered unacceptable, below a certain minimum threshold. Generally, growth profiles such as this one do not include these types of descriptions by fear of legitimizing inappropriate levels; this is certainly true of the Leithwood and Montgomery (1987) model, which describes growth in classroom innovation and professional development. In their model, the "typical" level includes only acceptable elements considered at or above the minimum competency threshold. This is due in part to the fact that the Leithwood model refers to personnel performance appraisal, in which individuals are duty-bound to perform according to certain behavioural norms. Given that, it is imperative that dereliction of duty is not represented or condoned in such a profile. In the present case, however, some elements that could be considered unacceptable or below the minimum competency standard have been included in order to properly reflect the reality of some government organizations. This is not to condone

such levels, but to provide the information necessary for some organizations to locate their current level of capacity, which may in fact be below certain acceptable standards.

Number of Levels

The original framework presented to participants included varying levels for each sub-dimension. The number of levels selected depended on the number of manageable steps required to move from low capacity to exemplary capacity, as per Leithwood and Montgomery (1987). This model offers significant flexibility in terms of allowing a more extensive discussion of the manageable steps required to move from one level to the next compared to other developmental models such as the seven-step framework of school improvement developed by Loucks-Horsley and Hergert (1985). However, while reading the framework, participants recommended the use of a standard number of levels for all of the sub-dimensions. Given this recommendation and the fact that the contents of each level within most sub-dimensions had to be reworded (more on this in the next section), the three levels (exemplary, intermediate and low capacity) were modified to remove additional levels while retaining the concept of manageability as organizations move from one level to the next. It is expected that this may change again as a result of the third phase of the study, which involves using the framework in an organizational context.

Location of Sub-Dimensions

Certain sub-dimensions were found to belong under a different dimension than had previously been thought. This was due in part to the revised content described below, and in part to reflect the differences between “capacity to do” evaluation and “capacity to use” evaluation. The sub-dimension on Organizational Infrastructure, previously part of the Organizational Characteristics dimension, was moved to Organizational Resources (dimension 2) because its revised contents focus on the other organizational units that support evaluation, such as HR and

procurement. In a similar vein, the Organizational Linkages sub-dimension was moved to Evaluation Planning and Activities (dimension 3) because the organizational relationships developed by evaluators are considered a part of their everyday activities. Finally, the Organizational Characteristics dimension was renamed “Evaluation Literacy” (dimension 4). This dimension retains Results-Management Orientation previously included under Organizational Characteristics, and also contains the former Evaluation Literacy sub-dimension, now renamed Involvement in Evaluation to avoid confusion.

Changes to Specific Dimensions

In addition to their broad comments on the features of the framework, participants provided detailed feedback on each dimension and its sub-dimensions. These comments as well as the changes made to the framework as a result of this feedback are presented in the following paragraphs.

Human Resources (dimension 1)

The Human Resources dimension elicited considerable feedback from the interview participants. There was general agreement about the choice of sub-dimensions and many suggestions were made to clarify the wording used throughout this section of the framework and to add to the elements already present.

A key element of the Staffing sub-dimension indicates that organizations with exemplary evaluation capacity have an appropriate balance of junior and senior evaluators. This notion of “appropriate balance” was deemed particularly important by participants, especially within the context of succession planning: “If you don’t have the recruitment strategy built in through the whole stream, then you’re not going to be able to achieve that balance” (Expert 3). An additional bullet was also added on developmental opportunities for evaluators. It was felt that an

exemplary organization would have an in-house development program for its staff, and that such a program would be key to evaluation capacity building. Finally, the reference to professional development was removed from this sub-dimension because of overlap with another sub-dimension in the same section of the framework.

The Evaluation Logic sub-dimension was revised based on the comments of one of the participants (Expert 2), who suggested that the “Evaluation Competency Profile” (Government of Canada, 2002) developed by the Treasury Board Secretariat (TBS) might be helpful in terms of wording, since it uses three levels similar to the ones used in the framework (junior, intermediate and senior). The elements included in this sub-dimension were therefore revised to include the contents of the “Cognitive Capacity” competency developed for the Evaluator Profile document.

The Communication and Interpersonal Skills sub-dimension was also modified based on the competencies identified for evaluators by TBS. The elements included in this sub-dimension were revised to include the contents of the “Communication” and “Interpersonal Relationships” competencies. Overall, participants agreed on the importance of this sub-dimension, stating that evaluators need presentation skills as well as the ability to understand other points of view “to summarize complex methods and findings effectively” (Expert 2) and to meet the informational needs of senior executives.

The adoption of the Treasury Board Evaluator Competency Profile to describe the levels of these sub-dimensions is based on the fact that a) an expert in government evaluation suggested its use, and b) the Competency Profile was developed as part of a study on evaluation capacity, conducted specifically for government by consultants.

Two changes were made to the Professional Development sub-dimension. First, the bullet on learning plans was removed due to concern from participants that this was too organization-specific to be of use in the framework. Second, a new bullet was added on internal professional development. Some participants felt that an organization with exemplary evaluation capacity would take the lead on developing its own internal capacity through professional development activities implemented in-house, such as brown bag sessions on evaluation, peer review processes for completed studies, and building a resource library for evaluators with books and journals.

The Leadership sub-dimension was found to be controversial among participants. On the one hand, participants recognized the importance of a good evaluation manager to increasing an organization's evaluation capacity; on the other, they had difficulty with the wording used to describe the three levels of leadership in the framework. For example, the issue of staff being comfortable with taking direction from the leader was thought to be unclear. This element was removed from all three levels of the sub-dimension. In an attempt to clarify the characteristics that define a leader in each of the three levels, the Evaluator Competency Profile was once again consulted; the elements included in the "Action Management" competency were integrated into the framework. Finally, the amount of experience in the "low capacity" level was changed to reflect the situation of individuals who are new to the area of evaluation, and a distinction was made between evaluation and management experience to better reflect the two separate skill sets. According to participants, an experienced evaluator does not necessarily make a good manager, and a competent manager does not necessarily always understand the complexities associated with research design and data analysis. The participants agreed that both facets were essential to

any evaluation unit leader: “the leader sits between technical experts and senior managers – they need to be able to translate from one to the other and get compromises when needed” (Expert 2).

Organizational Resources (dimension 2)

A number of changes were made to the three sub-dimensions included under Organizational Resources. These are outlined in the following paragraphs.

A few minor changes were made to the Budget sub-dimension to reflect the practices of government departments and agencies. First, the element on the approved multi-year evaluation plan and budget was removed because this was deemed to not be standard practice by participants. In most organizations, funding is only approved one year at a time, as per central agency requirements. Second, the words “adequate” and “sufficient” were removed in the exemplary and intermediate levels because participants felt that these had the same meaning and might lead to confusion. Finally, the element on the source of funding was changed to better reflect desired practice. Funding for evaluations should come from Treasury Board submissions rather than directly from the program’s budget: “programs are not supposed to fund evaluation; they are only supposed to ensure that separate funds are requested for evaluation and go directly to evaluation” (Expert 4).

Two main changes were made to the Ongoing Data Collection sub-dimension. First, a bullet was added on the successful implementation of performance measurement systems and the resulting quality of the data collected. According to participants, some organizations have good performance measurement systems, but they are not always well implemented, and as a result, data are collected sporadically or are missing altogether. Second, the scope of “evaluation studies” was broadened to include other uses of performance data; this may especially be the case in larger organizations that use performance data for a number of different purposes.

The newly renamed Evaluation Literacy sub-dimension, Involvement in Evaluation, was also modified slightly as part of the move from one dimension to another. The most important change made here is the removal of the participatory evaluation component previously included. Because of recent changes in the political landscape of federal government evaluation, a renewed focus on neutrality and objectivity has resulted in fewer participatory studies. Staff members are no longer encouraged to participate in all aspects of evaluation studies as they may have been in the past, but are still involved in certain evaluation activities:

...we (the evaluators) let them review it, but we're the technical experts and we're neutral and independent. So we don't want them developing data collection instruments, for example... in terms of identifying evaluation issues, we want to interview them, we would have a meeting and find out, but our job is to identify the issues and our job is to develop the instruments. That one in particular, because I wouldn't expect them to be able to do that anyways, although I also wouldn't see it as their role (Expert 3).

Furthermore, the client of the evaluation is now considered to be the Deputy Minister, or most senior individual in the organization. Because the program manager is no longer seen as the key client for the evaluation, there is less of a need to involve program staff in the evaluation process:

...for the evaluators, it's a real focus on relevance, value for money, credibility, DM is the client. That's really where... evaluation lacks credibility from the perspective of senior management and so you need to enhance the neutrality (Expert 3).

Therefore, this sub-dimension was modified to better reflect this new reality. Any mention of program staff involvement was changed to ensure that they are only involved in

certain parts of the process and the type of involvement changed from heavy involvement to offering feedback on certain aspects of the evaluation.

These changes highlight an important difference between the current policy direction taken by the federal government and existing knowledge and practices about the benefits of participatory evaluation revealed through evaluation research. Because the framework is meant to balance the input of experts against guidance from the evaluation literature, this creates somewhat of a conflict between the two views of stakeholder involvement. While it is not possible to reconcile these perspectives fully, an attempt was made to adequately represent the current thinking in government about this topic while including the potential for stakeholder participation by retaining program managers and staff in the descriptions provided for each level of capacity.

Finally, the Organizational Infrastructure sub-dimension was changed as it was moved into this dimension. New information was added to include more detail on the organizational supports that evaluation professionals need to complete their work, such as procurement, communications and HR and that are often considered, in the words of one participant, “organizational red tape” (Expert 3).

Evaluation Planning and Activities (dimension 3)

This dimension focuses on the actual practices of the evaluation unit and includes planning the work of the evaluation unit, establishing linkages with other organizational units, and conducting the evaluation work. Some key changes were made to the sub-dimensions included in this section of the framework in order to better describe the activities that actually take place in government departments and agencies.

The Evaluation Plan sub-dimension was modified slightly to include activities such as ongoing intelligence gathering and planning the evaluation unit's examination of its own service standards: "...it's not just reporting evaluations in the DPR and RPP, but reporting on evaluation itself" (Expert 4). References to approval of the evaluation plan were removed, since it is a mandatory activity that should not be influenced by the level of evaluation capacity of the organization. Some participants were concerned that leaving this bullet in the framework could be seen as something that's "easy to tick off" (Expert 1) and that would take away from the more important aspects of evaluation capacity.

A small change was made to the Use of Consultants sub-dimension to include a reference to interdepartmental evaluations as one of the potential uses of consultants: "sometimes just because it's interdepartmental it's awkward because you've got that outside body" (Expert 3).

The Information Sharing sub-dimension was modified in an attempt to clarify the wording of some of its elements. For example, the word "discuss" was changed to "share" to include a broader set of activities focused on dissemination of evaluation results and evaluation knowledge. An additional element was also included on policy and planning to ensure that the linkages between evaluation and policy were made explicit.

A number of elements were added to the External Supports sub-dimension to broaden its scope. For instance, the exemplary level was modified to include contacts with other organizations and academic experts, and all levels now include a reference to external stakeholders responsible for the delivery of federal programs:

It's important to consider external stakeholders as part of the federal government because they are often relied upon to deliver government programs. If they are not involved in the

evaluation process, they will not learn from it and then evaluation as a learning process is not sustainable (Expert 1).

Finally, the Organizational Linkages sub-dimension was moved into this section of the framework from the Organizational Decision-Making dimension because participants felt that evaluators should proactively seek out linkages with program personnel, senior management and other corporate functions such as policy development, audit and performance measurement. Contacts between evaluators and these groups should be considered routine in organizations that have exemplary evaluation capacity.

Organizational Characteristics (dimension 4)

The Results-Management Orientation sub-dimension was deemed to be critical to organizational evaluation capacity by participants. A few changes were made to clarify and add to the elements included under this sub-dimension. For example, the wording around the development of RMAFs was changed to highlight the fact that program managers are responsible for RMAFs – whether they choose to fulfill this responsibility varies in the three levels of capacity outlined in the framework. The term “logic modeling” was also changed to include other forms of results chains; according to one participant, anything “that depicts the relationships between outputs and outcomes” (Expert 1) should be included here. Finally, details were added about the enabling environment required to institute a results-management orientation in the organization. Participants felt that this environment should provide organizational members with the space and time that they need to learn about results-based management and to implement it properly.

Integration with Organizational Decision-Making (dimension 5)

In addition to the removal of the Organizational Linkages sub-dimension from this section of the framework and its inclusion in the Evaluation Planning and Activities dimension, a few changes were made to integrate comments made by interview participants.

The intermediate level of the Management Processes sub-dimension was modified slightly to clarify the wording used – instead of using the word “feedback”, participants felt that broadening the statement to include awareness and questions from organizational members about evaluation was necessary.

No major changes were made to the Decision Support sub-dimension, although some data from the interviews were used to describe each element more thoroughly. For example, participants stressed the need for clear evidence that evaluation supports organizational decisions and focused on the demand for evaluation services as a sign of exemplary evaluation capacity: “The defining thing is do they do it? Do they have a table at which they do it? Can you read the minutes of those meetings to find out if they are doing it?” (Expert 1).

Learning Benefits (dimension 6)

This dimension was found to have some overlap with other elements included in the framework, such as the Decision Support sub-dimension. The fact that learning takes place at an individual and organizational level as a result of evaluation is a pervasive theme throughout the framework and can be considered the ultimate objective of capacity building activities. Participants mainly commented on this dimension and made few suggestions for improvement. For example, they felt that evaluations “increase managers’ understanding of programs” (Expert 2), that evaluation requires a “continuous learning and enabling environment” (Expert 2) in order

to yield benefits to the organization, and that evaluation serves its purpose even when information that is already known is presented in a different way.

Other Comments on Framework

A number of comments were made by participants regarding the applicability of the framework and its potential usefulness in government organizations. Some participants felt that the framework could be of use to evaluation directors interested in assessing their evaluation capacity or as a guide when asking for specific resources for the evaluation unit. Another felt that the framework could be used as a maturity model for evaluation capacity, and that this model could guide central agencies in making decisions about the allocation of resources to the evaluation function. All participants thought that the framework could be modified in the future to be made into an instrument and applied across government. In order to do this, one participant suggested a detailed examination of documents produced by central agencies such as the Management Accountability Framework, the maturity model developed for modern comptrollership, a self-assessment tool developed for internal audit units and other organizational development models. These documents could provide useful information in terms of indicators, measurement scales, and would provide additional support for the framework's contents. These comments will be explored in more detail in the discussion section of the thesis.

Summary of Findings for Phase 2

This phase of the study focused on a second round of interviews with the same four participants as Phase 1. This time, participants were asked to review a draft framework of evaluation capacity and provide comments on its contents and format. Overall, participants felt that the format of the framework and its six dimensions were clear and comprehensive.

Suggestions were made concerning the specific wording of some elements and the location of some of the sub-dimensions.

A front page summary was included in the framework at the request of participants. Shaped similarly to a logic model, the summary shows all six dimensions and how they relate to one another, based on the evidence gathered in the interviews. The number of levels was also standardized throughout the framework based on the participants' comments, and the "typical or developing capacity" level was renamed to "low or developing capacity" in order to provide a more accurate picture of this level.

The Organizational Infrastructure sub-dimension was moved to Organizational Resources and now focuses more heavily on the organizational supports required by evaluators to complete their work effectively. The Organizational Linkages sub-dimension was moved to Evaluation Planning and Activities and features ongoing, regular contacts established by evaluators with other organizational units. Finally, the Organizational Characteristics dimension was renamed Evaluation Literacy and now includes the Results-Management Orientation and Involvement in Evaluation sub-dimensions.

A number of changes were made to the specific elements that comprise the sub-dimensions. These changes were based on participant comments and focus on clarifying the text as well as improving the contents of the framework.

Chapter 6 – Findings from Phase 3

This chapter summarizes the findings from the third phase of the study and provides a final version of the evaluation capacity framework. This phase of the study focused primarily on the validation of the framework in four federal government organizations with varying degrees of evaluation capacity. Detailed information on the process used in this phase of the study and the background of participants is provided in Chapter 3. The data collected through the eleven (11) interviews conducted in this phase were used to inform the changes made to the framework and to develop short summaries of evaluation capacity for each of the four participating organizations.

General Comments on Framework

Participants were asked about their general impression of the framework. Overall, they were fairly positive about both the contents of the framework and its presentation. Participants stated that the dimensions and sub-dimensions used were appropriate and felt that the organization of the dimensions in the “capacity to do” and “capacity to use” categories was helpful to their understanding. No major elements were identified as missing from the framework, although some additions were suggested for some of the sub-dimensions (these will be identified in the following section). Participants also commented on their experience with using the framework; in general, they felt that the wording was clear and that the framework gave them direction in terms of the various elements that they could implement in their organizations to increase evaluation capacity. In the words of one participant, “it helped jog my reflection on things that my organization doesn’t do” (Head of Evaluation 2).

One major issue raised by all of the participants is the difficulty they experienced in clearly differentiating between some of the levels. In particular, the difference between

intermediate and exemplary capacity was not always readily apparent to the participants. One participant explained this difficulty in terms of progression: “I couldn’t always differentiate between the levels because they weren’t always progressive from one level to the next; one doesn’t always build on the previous” (Head of Evaluation 2). Participants also felt that the “low or developing capacity” level would benefit from further differentiation, since developing capacity is not the same conceptually as low capacity. In their eyes, low capacity was more negative, while developing capacity represented some effort, albeit limited, at building an organization’s evaluation capacity: “In some cases, low implies negative. Is that a true reflection of low capacity? Low does not necessarily mean that they’re developing; they are missing the mark” (Evaluator 4). The fact that none of the four organizations identified themselves as having low or developing capacity on all dimensions may suggest a certain discomfort with admitting that the organization is not at all engaged in evaluation capacity building. Although none of the participants stated this outright, there may be little use in keeping the low dimension in future editions of the framework. However, because this has yet to be determined (perhaps in a future study related to the use of the framework in organizations), the low capacity level was not removed from the final version of the framework.

The dimensions and sub-dimensions included in the final version of the framework are the same as those presented in the Phase 2 draft. An additional level, however, was added in order to separate “developing” from “low” capacity. This was done for all dimensions and sub-dimensions, since participants felt that a standardized approach to the number of levels would make the framework clear and easy to use.

Figure 4: Summary of Phase 3 Framework

EVALUATION CAPACITY IN CANADIAN FEDERAL GOVERNMENT ORGANIZATIONS
 FRAMEWORK SUMMARY

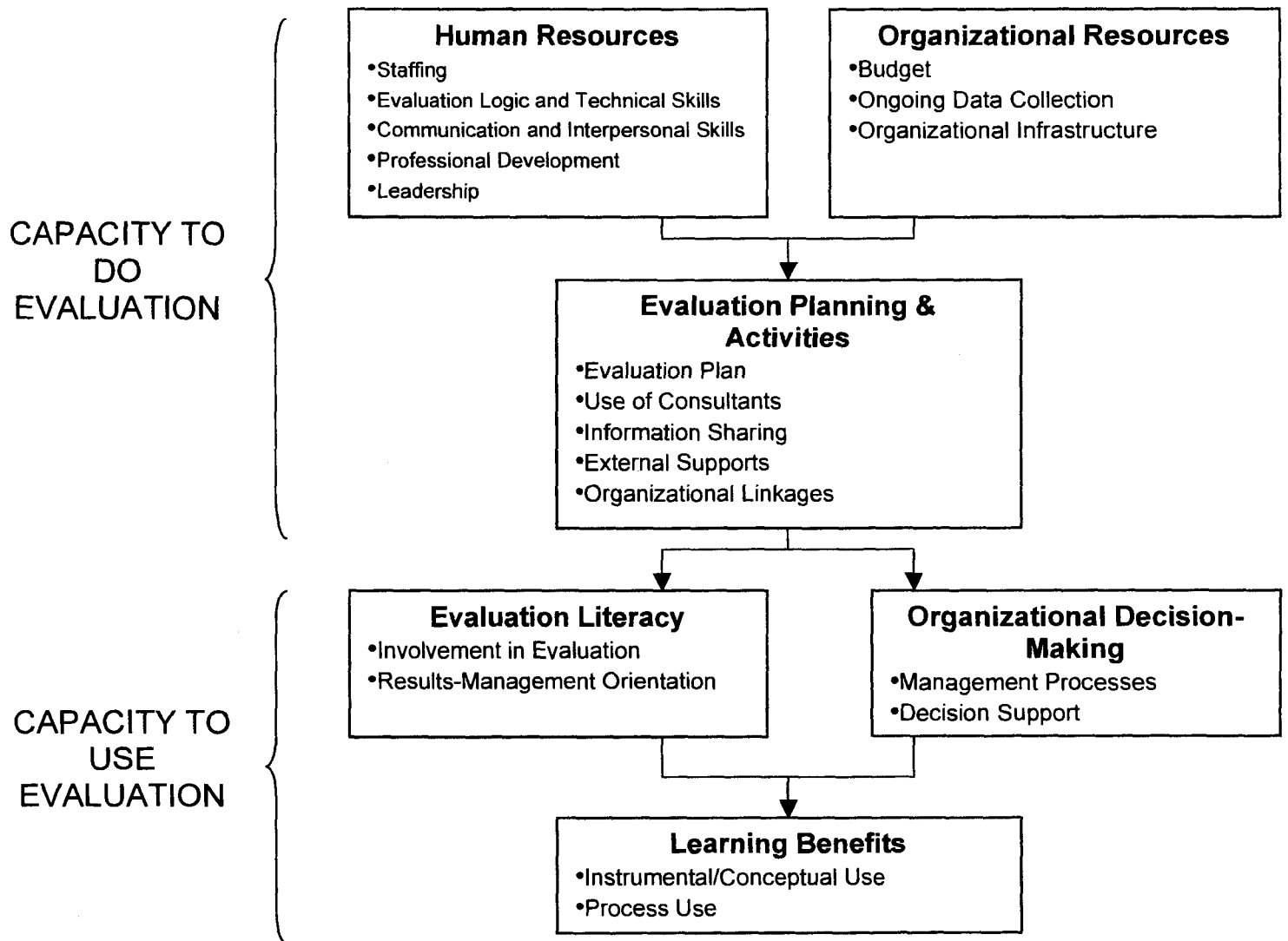


Table 5: Phase 3 Framework

CAPACITY TO DO EVALUATION

1. Human Resources					
Level	1.1 Staffing	1.2 Evaluation Logic and Technical Skills	1.3 Communications and Interpersonal Skills	1.4 Professional Development	1.5 Leadership
Exemplary Capacity	<ul style="list-style-type: none"> • Evaluation unit is optimally staffed (i.e., positions are created and filled based on operational requirements outlined in long-term evaluation plan) • Appropriate balance of senior and junior evaluator positions, given organizational requirements (i.e., size of organization, proportion of work done in-house, etc) • Career progression process is in place to facilitate promotions within unit and ensure succession planning through employee retention 	<ul style="list-style-type: none"> • Evaluation issues are clearly identified and linked to ongoing organizational concerns and priorities • Innovative use of methods and approaches to data collection (e.g., routine use of complex survey methodology) • Recommendations made in evaluation reports are clearly linked to evaluation findings • Evaluation projects are generally well-managed and problems are usually identified and resolved quickly by senior evaluators 	<ul style="list-style-type: none"> • Evaluation unit has clearly established client trust within the organization • Evaluation reports and other products deliver open, clear and transparent messages (i.e., are readily accepted by clients) • Clients feel that evaluators understand key organizational issues and respond to them appropriately (i.e., informational needs of program managers are met through various evaluation activities) 	<ul style="list-style-type: none"> • Assessment of skill sets among staff is done regularly and learning activities are arranged to fill gaps within unit • All team members have personalized learning plans • All staff members engage in external professional development activities directly related to their work (e.g., conferences) • Organization develops its own internal professional development activities (e.g., brown bag sessions, resources, in-house seminars, etc) 	<ul style="list-style-type: none"> • Evaluation unit headed by individual with strong evaluation and management background • Leader effectively reconciles expectations of senior management with operational requirements and resources of team • Leader guides, mentors or coaches team members as part of their regular duties

<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> • Evaluation unit is fully staffed (i.e., all positions are filled and adequate to meet operational requirements of annual evaluation plan) • Appropriate balance of senior and junior evaluator positions, given organizational requirements (i.e., size of organization, proportion of work done in-house, etc) • Career progression process is available for some levels of evaluation professionals wishing to gain experience 	<ul style="list-style-type: none"> • Evaluation issues are clearly identified and reflect the concerns of program managers • Use of evaluation methods respects accepted standards and yields defensible findings (i.e., well-developed interview protocols) • Recommendations made in evaluation reports are clearly linked to evaluation findings • Evaluation projects are fairly well-managed and problems are usually identified and resolved with help from evaluation unit leader 	<ul style="list-style-type: none"> • Evaluation unit is committed to building and maintaining client trust • Evaluation reports deliver open, clear and transparent messages (i.e., are usually accepted by clients) • Clients feel that evaluators understand key organizational issues (i.e., informational needs of program managers are taken into account in evaluation design) 	<ul style="list-style-type: none"> • Most team members have personalized learning plans • Most staff members engage in external professional development activities directly related to their work (i.e., conferences) • Some internal learning resources are made available to staff members (e.g., journals, books) 	<ul style="list-style-type: none"> • Unit headed by individual with some evaluation and management experience • Leader effectively reconciles expectations of senior management with operational requirements and resources of team • Leader guides teams members as often as possible
<p>Developing Capacity</p>	<ul style="list-style-type: none"> • Evaluation unit is partially staffed (i.e., less than 50%; requirements outlined in annual evaluation plan) • Few senior evaluation positions in the unit • Career progression occurs in an ad hoc manner (i.e., through the usual competition process when turnover occurs) • No succession planning efforts are underway 	<ul style="list-style-type: none"> • Evaluation issues are clearly identified • Use of evaluation methods respects accepted standards and yields defensible findings (i.e., well-developed interview protocols) • Recommendations made in evaluation reports are usually linked to evaluation findings • Some project management issues may arise due to lack of experience or resources 	<ul style="list-style-type: none"> • Evaluation unit is working towards building client trust • Evaluation reports are generally written in a clear manner (i.e., clients suggest some edits to the reports or ask some clarification questions) • Clients feel that evaluators are open to learning about issues related to program management (i.e., program managers are able to make suggestions regarding evaluation questions) 	<ul style="list-style-type: none"> • Some staff members engage in professional development activities delivered by external organizations to improve generic skill sets (i.e., not necessarily directly related to evaluation) 	<ul style="list-style-type: none"> • Unit headed by individual who is new to the area of evaluation and/or has limited management experience • Leader is not generally involved in senior management discussions and therefore assigns work based only on operational or TB requirements • Leader coordinates team activities but is not involved in guiding team members in their work
<p>Low Capacity</p>	<ul style="list-style-type: none"> • Evaluation unit has several vacant positions (i.e., more than 50%; no link between evaluation plan and staffing actions) • No senior evaluator positions in the unit • No career progression process for evaluators due 	<ul style="list-style-type: none"> • Evaluation issues are not always identified • Weak evaluation methods do not always yield defensible findings (i.e., reliance on external data sources with no verification) • Recommendations made in evaluation reports are not 	<ul style="list-style-type: none"> • Evaluation unit has not yet been able to develop client trust • Evaluation reports often raise clarification questions from clients (i.e., clients do not understand the chain of results or specific pieces of evidence) • Clients feel that evaluators are not open to learning about issues 	<ul style="list-style-type: none"> • Staff members do not engage in professional development activities delivered by external organizations or provided in-house 	<ul style="list-style-type: none"> • Head of evaluation position is vacant

	to small number of available positions in organizational chart	<ul style="list-style-type: none"> No succession planning efforts are underway 	inked to evaluation findings	<ul style="list-style-type: none"> Problems with project management process arise often due to lack of experience or resources 	related to program management (i.e., program managers are not able to make suggestions regarding evaluation questions)	
--	--	---	------------------------------	---	--	--

CAPACITY TO DO EVALUATION

2. Organizational Resources		2.3 Organizational Infrastructure	
Level	2.1 Budget	2.2 Ongoing Data Collection	2.3 Organizational Infrastructure
Exemplary Capacity	<ul style="list-style-type: none"> Evaluation budget ensured through continuing funding specifically allocated to evaluation unit (i.e., is not shared with other corporate units) Evaluation budget is allocated based on evaluation plan (i.e., specific amount of budget based on plan) 	<ul style="list-style-type: none"> Performance measurement system fully implemented across organization (or across programs in the case of large organizations) High-quality information is collected by the PM system Performance data feeds directly into results-based management (including evaluation studies) 	<ul style="list-style-type: none"> The organization has a stable governance structure and clear accountability lines for results Organizational policies on evaluation and performance measurement have been developed and implemented Organizational culture ensures that performance data feed into a structured planning and reporting process (e.g., use of evaluation in RPP, budgeting, etc) Existence of organizational supports that provide needed services to evaluators in a competent and timely manner, allowing them in turn to produce timely and useful evaluation reports (i.e., procurement, communications, HR, access to information etc) Organizational infrastructure showing some maturity (e.g., clear governance structure and stability in senior management ranks; organizational policy on evaluation has been developed and implemented) Existence of a structured planning process that includes consideration of evaluation issues and findings (e.g., use of evaluation in RPP, budgeting, etc) Organizational supports do not always provide needed services to evaluators in a timely manner and results in delays for evaluation projects
Intermediate Capacity	<ul style="list-style-type: none"> Evaluation budget is stable but shared with some other corporate groups (i.e., policy or audit) Evaluation budget is appropriate given the evaluation plan (i.e., plan does not determine the budget, but it is sufficient to complete planned activities) 	<ul style="list-style-type: none"> Performance measurement system in place for major programs or activities Incomplete implementation of PM system (i.e., sporadic data collection, missing variables, etc) Performance data can be adapted to suit the informational requirements of results-based management (including evaluation studies) 	

<p>Developing Capacity</p>	<ul style="list-style-type: none"> • Evaluation budget provided on a case-by-case basis for each new project • Evaluation plan not considered in budget allocation 	<ul style="list-style-type: none"> • Performance measurement is done on a program-by-program basis • Ad hoc implementation of performance measures, with uneven quality • Performance data is difficult to integrate into results-based management (including evaluation studies) • Performance measurement is not done in the organization, either by programs or other corporate groups 	<ul style="list-style-type: none"> • Organizational infrastructure still under development (e.g., no clear governance structure, no policies on evaluation are in place, no structured planning process) • Organizational supports often fail to provide needed services to evaluators
<p>Low Capacity</p>	<ul style="list-style-type: none"> • Organizational budget does not include funds for evaluation activities 		<ul style="list-style-type: none"> • Organizational infrastructure still under development (e.g., no clear governance structure, no policies on evaluation are in place, no structured planning process) • Evaluators do not have regular access to organizational supports

CAPACITY TO DO EVALUATION

3. Evaluation Planning and Activities	
Level	3.5 Organizational Linkages
Exemplary Capacity	<p>3.1 Evaluation Plan</p> <ul style="list-style-type: none"> • Evaluation plan follows a five-year cycle and is updated annually • Evaluation plan developed in consultation with all senior managers and includes needs assessment exercise • Evaluation plan includes thorough risk assessment process • Ongoing intelligence gathering allows for changes to be made to plan when necessary (i.e., integrate emerging needs) • Evaluation plan includes systematic review of evaluation unit itself (i.e., establishing and measuring service standards for the unit, assessing impact on organization)
	<p>3.2 Use of Consultants</p> <ul style="list-style-type: none"> • Appropriate balance of evaluations designed and conducted by evaluation staff, or by consultants for specific expertise (e.g., when a project's scope is too large for the evaluation unit's resources, when dealing with complex interdepartmental evaluations or when specialized expertise is required) • High-quality work produced by consultants when they are involved
	<p>3.3 Information Sharing</p> <ul style="list-style-type: none"> • Major decisions on evaluation projects are discussed within the unit to benefit from staff members' knowledge and experience • Evaluators actively gather information on new developments in policy and strategic planning • Knowledge management issues and processes are discussed regularly and common standards are followed by staff members
	<p>3.4 External Supports</p> <ul style="list-style-type: none"> • Evaluators make frequent use of external supports to evaluation such as professional association, published quality standards, etc. • Evaluators are actively involved in broadening their external networks by liaising with evaluators in other organizations, engaging academic experts, and gathering information on activities of OAG, TBS, Privy Council, etc • External stakeholders responsible for the delivery of federal government programs are involved in evaluation activities (where applicable)
	<p>3.5 Organizational Linkages</p> <ul style="list-style-type: none"> • Evaluators in regular contact with program managers through formal or informal ties • Ready access to ADM and DM on all aspects of evaluation; clear interest in evaluation information demonstrated by ADM and DM • Evaluation unit located in close proximity to key organizational areas such as policy development, strategic planning and performance measurement units

<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> • Evaluation plan follows a five-year cycle and is updated annually • Evaluation plan developed in consultation with most senior managers • Evaluation plan includes some assessment of risk • Provisions made for reviewing the plan on an ongoing basis 	<ul style="list-style-type: none"> • Medium and large evaluations are designed in-house and conducted by external consultants • Evaluation staff involved in conducting the evaluation (i.e., contribute to some of the field work done by consultants) • Smaller studies are conducted in-house by evaluation staff • Consulting work produced is generally considered high-quality 	<ul style="list-style-type: none"> • Evaluators share their progress and other information with their colleagues at regular staff meetings • Evaluators are generally aware of new developments in policy and strategic planning • Knowledge management standards have been developed and are generally followed within the unit 	<ul style="list-style-type: none"> • Evaluators make some use of external supports to evaluation such as professional association, published quality standards, etc. • Evaluators keep themselves informed of new developments in external organizations of interest (i.e., universities, other departments and agencies, central agencies, etc) • External stakeholders involved in the delivery of federal government programs are kept informed 	<ul style="list-style-type: none"> • Evaluators communicate with their program clients through ongoing, formal mechanisms • DM and ADM receive regular reports about evaluation activities but are not directly involved in evaluation • Evaluation unit located in close proximity to some key organizational areas such as policy development, strategic planning and performance measurement
<p>Developing Capacity</p>	<ul style="list-style-type: none"> • Evaluation plan exists for one or two years • Evaluation plan developed in consultation with some program managers or senior managers • No consideration of risk in planning evaluation schedule 	<ul style="list-style-type: none"> • Virtually all evaluations are designed and conducted by external consultants • Evaluation staff is mainly involved in managing contracts and overseeing the work done with little substantive input • Varying quality levels in work produced by consultants 	<ul style="list-style-type: none"> • Evaluators share their progress with their supervisors and other staff members in a sporadic, informal manner • Evaluators are not aware of new developments in policy and strategic planning • Knowledge management standards exist but are not followed 	<ul style="list-style-type: none"> • Evaluators have access to basic external supports, such as a professional association or published quality standards but do not often make use of them • Evaluators do not generally liaise with external organizations or experts • External stakeholders are informed of evaluation reports once they are published (where applicable) 	<ul style="list-style-type: none"> • Evaluators communicate with program clients on specific issues related to projects • DM and ADM are made aware of evaluation findings only through formal requirements (e.g., internal evaluation committee) • Evaluation unit is removed (either physically or structurally) from key organizational areas such as policy development, strategic planning and performance measurement
<p>Low Capacity</p>	<ul style="list-style-type: none"> • No evaluation plan is developed by staff • Evaluation projects occur on ad hoc basis 	<ul style="list-style-type: none"> • All evaluation work is contracted out • Evaluation staff manage contract work done by consultants with no substantive input • Some problems with the quality of the work produced by consultants 	<ul style="list-style-type: none"> • Evaluators do not typically share progress with other staff members • Evaluators are not aware of new developments in policy and strategic planning • Knowledge management standards do not exist within the unit 	<ul style="list-style-type: none"> • Evaluators do not have access to basic external supports, such as a professional association • Evaluators do not liaise with external organizations or experts • External stakeholders are not informed of evaluation 	<ul style="list-style-type: none"> • Evaluators communicate infrequently with program clients • DM and ADM tend to delegate responsibility for evaluation • Evaluation unit is removed (either physically or structurally) from key

CAPACITY TO USE EVALUATION

Level	4.1 Involvement in Evaluation	4.2 Results-Management Orientation
<p>Exemplary Capacity</p>	<p>4.1 Involvement in Evaluation</p> <ul style="list-style-type: none"> Organizational staff members generally understand the purpose of evaluation and how it supports the organizational mandate (e.g., staff members understand results-based management principles and practices) Program managers and other staff members are closely involved at key points in the evaluation process (e.g., review identified issues and provide feedback, facilitate data collection opportunities, review draft evaluation reports, etc) 	<ul style="list-style-type: none"> Senior managers promote a results-management orientation for the entire organization and make it a priority by providing time and resources Organizational members share clear ideas about organizational purpose and goals through formal and informal mechanisms (e.g., strategic planning sessions, retreats, regular meetings, brown bag lunch sessions, etc) All programs have a clear results chain (i.e., logic model) Program managers take the lead for the development and implementation of RMAFs and other performance management activities; evaluators provide technical expertise when needed Organizational outcomes or expected results are only outlined in official documentation but are not included in communications from senior managers Organizational members share clear ideas about organizational purpose and goals through formal mechanisms (i.e., strategic planning sessions, meetings, etc) Some programs have a clear results chain (i.e., logic model) Program managers work with evaluators in the development and implementation of RMAFs and other performance measurement activities, but evaluators lead these projects
<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> Organizational staff members are familiar with the general principles of evaluation and how it can help them in their work (e.g., they understand the difference between evaluation and audit) Program managers are involved in evaluation projects (e.g., sit on Evaluation Steering or Advisory Committees) and provide program-related feedback on report drafts 	<ul style="list-style-type: none"> Organizational outcomes or expected results are not articulated clearly for all organizational members; most are not aware of results management principles and practices Some programs are engaged in developing results chains such as logic models Program managers not involved in the development or implementation of RMAFs and other performance management activities; evaluators conduct these processes with little input from programs Organizational outcomes or expected results have not been developed Programs do not have results chains such as logic models The organization does not support the development of RMAFs or other performance measurement activities
<p>Developing Capacity</p>	<ul style="list-style-type: none"> Little awareness of evaluation or its purpose within larger organizational context Little involvement from program staff and managers (i.e., brief comments on draft evaluation reports) 	<ul style="list-style-type: none"> Organizational outcomes or expected results are not articulated clearly for all organizational members; most are not aware of results management principles and practices Some programs are engaged in developing results chains such as logic models Program managers not involved in the development or implementation of RMAFs and other performance management activities; evaluators conduct these processes with little input from programs Organizational outcomes or expected results have not been developed Programs do not have results chains such as logic models The organization does not support the development of RMAFs or other performance measurement activities
<p>Low Capacity</p>	<ul style="list-style-type: none"> No discernible awareness of evaluation or its purpose within larger organizational context No involvement from program staff and managers 	<ul style="list-style-type: none"> Organizational outcomes or expected results are not articulated clearly for all organizational members; most are not aware of results management principles and practices Some programs are engaged in developing results chains such as logic models Program managers not involved in the development or implementation of RMAFs and other performance management activities; evaluators conduct these processes with little input from programs Organizational outcomes or expected results have not been developed Programs do not have results chains such as logic models The organization does not support the development of RMAFs or other performance measurement activities

CAPACITY TO USE EVALUATION

5. Integration with Organizational Decision-Making	
Level	5.2 Decision Support
Exemplary Capacity	<p>5.1 Management Processes</p> <ul style="list-style-type: none"> Program and policy staff integrate evaluation into other areas of their work (e.g., they routinely request the involvement of evaluators in management processes such as MC and TB submission preparation)
Intermediate Capacity	<ul style="list-style-type: none"> Program and policy staff are aware of the evaluation services that can be provided and sometimes contact evaluation staff for advice
Developing Capacity	<ul style="list-style-type: none"> Evaluation unit operates separately from program units and is not generally involved in management processes; program and policy staff unaware of the potential contributions of evaluation staff
Low Capacity	<ul style="list-style-type: none"> Evaluation unit does not involve or inform program units of its activities
	<ul style="list-style-type: none"> Evaluation findings and recommendations considered in budget allocation and other high-level organizational and policy decisions Demand for evaluation services originates from all levels of the organization Evaluation findings and recommendations usually considered in program management decisions and some policy decisions Program managers are interested in and use evaluation as a management support tool (i.e., evaluation as provider of ongoing management information) Little consideration of evaluation findings and recommendations in organizational and policy decisions No specific demand for evaluation services other than to meet TB requirements Evaluation findings and recommendations are not used in organizational and policy decisions No demand for evaluation services exists within the organization

CAPACITY TO USE EVALUATION

6. Learning Benefits	
Level	6.2 Process Use
<p>Exemplary Capacity</p>	<p>6.1 Instrumental/Conceptual Use</p> <ul style="list-style-type: none"> Evaluation findings are used consistently as a basis for action and change (i.e., evaluation recommendations are appropriate and implemented in a timely manner) Evaluation findings and reports often have an impact on stakeholders' understanding and attitudes about programs
<p>Intermediate Capacity</p>	<ul style="list-style-type: none"> Evaluation findings are sometimes used as a basis for action and change (i.e., evaluation recommendations are sometimes implemented) Evaluation findings and reports can have an impact on stakeholders' understanding and attitudes about programs
<p>Developing Capacity</p>	<ul style="list-style-type: none"> Evaluation findings are rarely used as a basis for action and change (i.e., evaluation recommendations are usually not implemented) Evaluation findings and reports rarely have an impact on stakeholders' understanding and attitudes about programs
<p>Low Capacity</p>	<ul style="list-style-type: none"> Evaluation findings are never used as a basis for action and change (i.e., evaluation recommendations do not usually make their way to those with the ability to act upon them) Evaluation findings and reports do not have an impact on stakeholders' understanding and attitudes about programs (because they are rarely aware of the evaluation)

Comments on Sub-Dimensions

The bulk of the interview time in this phase of the study was spent going over the framework with each participant and recording his or her thoughts on the individual elements making up each of the sub-dimensions. These comments were coded by sub-dimension and by type (wording issue, differentiation between levels, suggested additions, other comments). The coded data were then used as the basis for the changes made to the framework. These changes are described in detail in the pages that follow.

Sub-dimension 1.1: Staffing

The staffing sub-dimension generated a number of comments from Heads of Evaluation as well as senior evaluators. The notion of a fully staffed evaluation unit, in particular, was identified as an area of concern given current difficulties in attracting and retaining evaluation professionals. In addition to this, a fully staffed unit did not equate with exemplary capacity in participants' minds. To some participants, "the notion of fully staffed depends on the evaluation plan, which in turn depends on the budget; we have to ask if the plan is realistic given the budget" (Evaluator 3). Full staffing was therefore changed to "optimal" staffing at the exemplary level, which refers to the ability of the evaluation unit to deliver based on the evaluation plan. In other words, exemplary organizations are able to create and modify positions to ensure that the program activities of the organization are adequately evaluated given the guidelines provided by Treasury Board. This is different from the intermediate level, in which "full staffing" refers to the fact that all positions are filled and are adequate for completing the work identified in the evaluation plan. The element of flexibility found at the exemplary level is missing from the intermediate level.

Another issue raised by participants refers to the “appropriate balance” of senior and junior evaluators identified at the exemplary level. Some participants asked for further clarification on the specific meaning of “appropriate”, while others suggested that the balance of senior and junior evaluator positions was less important than ensuring that the evaluation unit included the right skill sets for the work required. To these participants, the junior and senior evaluator positions don’t always match the experience and knowledge of those who fill them. In this context, it is more important to ensure that the individuals in place are able to do the evaluation work rather than focus on seniority issues: “the balance of junior and senior evaluators depends on whether the department does in-house evaluation, it’s more about the skill sets you have” (Evaluator 3). The issue of skill sets will be examined further in sub-dimension 1.2; in terms of staffing, the framework now focuses more specifically on the distribution of human resources than on the characteristics of the individuals currently employed.

Finally, some participants suggested the inclusion of a reference to succession planning. This was incorporated to the element dealing with career progression, since these two concepts are often linked in HR management and planning processes.

Sub-dimension 1.2: Evaluation Logic and Technical Skills

Although the descriptions for this sub-dimension were based on the Evaluator Competency Profile developed by Treasury Board, the fact that they refer to individual characteristics rather than organizational features made it difficult for participants to differentiate between the three levels of capacity. In fact, exemplary organizations should, according to participants, include evaluators with skills reflected by each of the three levels. In other words, junior evaluators would have the skills identified in the “low or developing capacity” level, more experienced evaluators would have the skills identified in the intermediate category, and senior

evaluators would have the skills found in the exemplary capacity level. It cannot be expected, therefore, that all staff members would have the skills found at the exemplary capacity level.

To improve the clarity of this sub-dimension, participants suggested changing the elements to focus on organizational characteristics rather than on individual skills. This, they argued, would have the benefit of placing this sub-dimension on the same scale as the others and provide a better comparison point for those using the framework. New elements were therefore developed to replace the ones found in the draft framework: using evaluation methods appropriately, producing clear and credible reports that include evidence-based recommendations, and managing projects effectively. Instead of focusing on the individual, these elements now describe the skills found within the evaluation unit.

Sub-dimension 1.3: Communications and Interpersonal Skills

This sub-dimension was also developed based on the Evaluator Competency Profile. Therefore, its elements were changed to reflect a more organizational focus rather than a description of individual characteristics. More specifically, this sub-dimension now includes elements focusing on establishing client trust, producing clear and transparent reports, and understanding the organizational context in which evaluation operates. These also incorporate other aspects of communications not previously included in the framework, such as written communications and listening skills crucial to successful evaluation.

Sub-dimension 1.4: Professional Development

Participants were generally pleased to see that professional development had been included as a sub-dimension of evaluation capacity and had little feedback on the various elements that describe it. The notion of peer review was removed from the framework because participants felt that this deals more with quality assurance than professional development. The

differentiation between the exemplary and intermediate levels was also made clearer by including an element on learning plans and on quantifying the element related to external professional development (i.e., all staff vs. most of the staff).

Sub-dimension 1.5: Leadership

A few changes were made to this sub-dimension to clarify the wording used and to better differentiate between the exemplary and intermediate levels. The second element (bullet), in particular, required some reworking in order to make its meaning clearer. As one participant described, evaluation leaders need to “manage up and down at the same time” (Evaluator 1). In other words, they must establish a bridge between the informational needs of senior managers and the project planning that occurs within the evaluation unit. This is part of the planning process considered in sub-dimension 3.1, but it is also an indicator of exemplary leadership. A noteworthy comment was made by one participant regarding the upcoming certification requirement for Heads of Evaluation that will likely be included in the new Treasury Board Evaluation Policy. Although this is not currently in place, such a requirement should be integrated to the framework if and when this becomes part of the policy.

Sub-dimension 2.1: Budget

As in previous phases, this sub-dimension garnered a number of comments from participants. Again, these comments focused on clarification and on improving the differentiation between the exemplary and intermediate levels. For example, participants were unclear about the difference between the continuing (A-base) funding mentioned in the exemplary level and the stable funding sources mentioned in the intermediate level. As one participant described, these definitions do not adequately represent what goes on in most organizations: “We receive a small A-base, then scrounge from programs and Treasury Board

submissions for money for our evaluations” (Head of Evaluation 2). These elements were changed to highlight continuing funding specifically allocated to the evaluation unit in the case of exemplary capacity, and funding allocated to a larger corporate group within which the evaluation unit resides in the case of intermediate capacity.

A second element was added to further differentiate between the two levels. In exemplary organizations, the evaluation budget is allocated based on the evaluation plan. In other words, the evaluation priorities are established first, and a budget is then provided to meet the operational requirements of the plan. In organizations situated at the intermediate level, the evaluation budget is adequate to complete the planned activities, but it is not allocated on the basis of the evaluation plan alone. It may be, for example, allocated based on available funding for the year or on the previous year’s budget.

Sub-dimension 2.2: Ongoing Data Collection

The opinions of participants varied on this sub-dimension, depending on whether they were situated in a large department or a small agency. For example, the presence of multiple performance measurement systems may be considered exemplary in larger organizations, while this would not be acceptable in smaller organizations where the streamlining of data collection is possible and advisable.

Sub-dimension 2.3: Organizational Infrastructure

Participants felt that this sub-dimension demonstrated good progression between the levels. A wording change was suggested for the third element of the exemplary level, to replace “mechanisms” by organizational culture. In the words of one participant, feeding performance data into broader reporting systems “needs to be internalized; managers need to look at results information regularly” (Head of Evaluation 2).

Some of the participants felt that the existence of an evaluation policy does not necessarily reflect exemplary capacity and that the Treasury Board Evaluation Policy offers adequate direction to departments and agencies. However, they did acknowledge that the development of an organizational evaluation policy may provide an opportunity to highlight roles and responsibilities and educate other organizational members about their place in the evaluation process. Therefore, no change was made on this point in the framework.

Sub-dimension 3.1: Evaluation Planning and Activities

The differentiation between the exemplary and intermediate levels was identified as a significant problem in this sub-dimension. The references made to “multi-year” and to “five-year” plans, in particular, resulted in some confusion for participants. Because five years is the standard length of time for most evaluation plans (as per Treasury Board policy), this was included in both the exemplary and intermediate levels. Even though “there may be valid reasons why a plan wasn’t made for five years” (Evaluator 3), an evaluation plan spanning less than five years was classified under developing capacity and low capacity, since this may indicate a broader capacity problem within the organization.

The main differences between the exemplary and intermediate levels are qualitative in nature and largely refer to the process used to develop and implement the evaluation plan. For example, at the exemplary level, the evaluation plan is developed in consultation with all senior managers and includes a needs assessment component. At the intermediate level, the evaluation plan may include consultation of some senior managers and does not include needs assessment. Exemplary capacity can also be demonstrated by ongoing intelligence gathering and by integrating emerging needs into the plan. Finally, exemplary organizations engage in systematic

reviews of the evaluation unit itself to provide an example to other groups. This appears to be a fairly rare occurrence in the Canadian federal government at present.

Sub-dimension 3.2: Use of Consultants

This sub-dimension also raised a number of comments from participants. Some questioned the inclusion of this sub-dimension outright, stating that the use or non-use of consultants is “not necessarily a determination of evaluation capacity” (Evaluator 4); rather, the amount of substantive input provided by evaluators within the context of outsourced projects is key. In other words, the true definition of capacity touches on the involvement of evaluators either as full partners in the project or as contract managers. This is reflected in the element focusing on the “appropriate balance” between in-house work and outsourcing included in the exemplary capacity level.

Another suggestion made by participants focused on the quality of the products submitted by consultants. Indeed, consulting work of poor quality has a direct impact on the credibility of internal evaluators and contributes to a lack of use of evaluation results. It was felt that, in some cases, this was a direct result of internal capacity issues, and therefore, this element has been added to the framework. For instance, this speaks to both the technical skills of the evaluators who manage the project (i.e., their ability to recognize quality issues and to raise them at the appropriate times) and to the process put in place by organizations to access consultant resources (i.e., standing offers, use of temporary staff, etc).

Sub-dimension 3.3: Information Sharing

A number of wording issues were raised for this sub-dimension. For example, participants were unclear about the difference between unit and team. This has been clarified wherever possible. In addition to this, the word “all” was removed in the first element (bullet)

after one participant expressed concern over the inclusion of all evaluators in decision-making processes: “I can’t really do that because we have such a large team. It would be impossible to involve everyone” (Head of Evaluation 1). Another participant suggested that decisions could be discussed within the evaluation unit rather than made by the group: decisions would be made by individuals, but “informed by the council of others” (Evaluator 4).

A third element was added on knowledge management issues and processes. According to some participants, a significant amount of time is spent discussing issues such as file storage, report names, and other practical processes. This is not necessarily part of specific projects, but it does represent an important portion of the information that is shared within the unit.

Sub-dimension 3.4: External Supports

Few comments were made about this sub-dimension. One participant suggested that an additional level was needed between the low capacity level and the intermediate level. This was done across the framework. Some participants questioned the third element dealing with external stakeholders delivering government programs. In the words of one participant, “it’s a smart thing to do in an evaluation but I don’t know if it’s really a support of evaluation” (Head of Evaluation 2). This is a situation particular to a small number of departments and agencies, and was deemed satisfactory by participants working in these organizations. Therefore, the element was not removed but modified slightly to clarify its meaning.

Sub-dimension 3.5: Organizational Linkages

Overall, participants felt that this sub-dimension was clear and that the differentiation between the levels was adequate. One participant questioned the need for organizational linkages and the impact that these may have on the perceived independence of the evaluators: “What about our independence as evaluators?” (Head of Evaluation 4). In fact, evaluators use their

organizational linkages to provide better services and produce results that are relevant and appropriate given the needs of their clients. This does not mean that their independence is compromised, and so no changes were made as a result.

Sub-dimension 4.1: Involvement in Evaluation

A few changes were made to this sub-dimension to improve the clarity of the wording and to differentiate between the levels more easily. For example, evaluation steering committees now also include advisory committees, since either type of committee can be created at the outset of an evaluation, depending on organizational tradition and preference. These were also made into examples, since not all departments have those committees. Other examples were also added to existing elements to clarify certain points. Although participants felt that this particular sub-dimension was “a nice concept” (Head of Evaluation 2), most thought that an exemplary level would be difficult to reach given current Treasury Board guidelines on evaluation. Nonetheless, the sub-dimension was retained in the framework based on the results of the in-depth literature and because participants agreed in principle that this was an important component of organizational evaluation capacity.

Sub-dimension 4.2: Results-Management Orientation

Very few comments were provided by participants for this sub-dimension, although there was general agreement about its usefulness in the framework. In addition to developing the RMAF, participants suggested adding its implementation as a key feature of results-based management. The responsibility for leading the RMAF development and implementation was also revised based on participant comments.

Sub-dimension 5.1: Management Processes

Participants felt that this sub-dimension was particularly important in terms of evaluation use. Many of them suggested, however, that further detail was necessary to differentiate between the exemplary and intermediate levels. Therefore, the two descriptions were changed to reflect a more obvious qualitative difference between them.

Sub-dimension 5.2: Decision Support

Although participants lamented the fact that “evaluation clients do not always implement the recommendations made by evaluators” (Head of Evaluation 2), most felt that the elements were appropriate and recommended no changes.

Sub-dimensions 6.1: and 6.2 Instrumental/Conceptual Use and Process Use

These sub-dimensions have been grouped together here because very few comments were made by participants on these, and the comments that were made indicated some confusion between the definitions of instrumental, conceptual and process use. One interesting comment made by a number of evaluation professionals focused on process use and the application of evaluation logic to other areas of a stakeholder’s work. Evaluation professionals are not necessarily aware of this kind of impact, unless it is specifically studied within the organization. In other words, a program manager who uses the general principles of evaluative inquiry before making a decision about program delivery won’t necessarily tell the evaluator about their thought process and how it may have been influenced by their recent experience with evaluation.

Summary of Findings on Framework

Overall, the participants felt that the dimensions and sub-dimensions were reflective of their own definition of evaluation capacity and did not find that any of these were missing. The comments made by participants on the framework can generally be placed in one of three

categories: wording clarification, differentiation between levels, and suggested additions to some of the specific elements included in the framework. Most of the sub-dimensions have been modified according to the participants' comments, and as a result, the final framework contains more information and more detail than previous drafts. An interesting observation, which will be considered at greater length in the discussion section, is the strong focus put by participants on the "capacity to do" dimensions. Because two-thirds of the interviewees were evaluation professionals, the findings reflect a greater consideration of the practical aspects of evaluation capacity, such as staffing and evaluation planning. These are dimensions that are more readily under the control of evaluators, and thus it appears as though they have received more thought than the more difficult "capacity to use" dimensions, which often depend on a number of external factors. In terms of use, however, the "capacity to use" dimensions are more important and demonstrate a higher level of evaluation capacity than the dimensions included under "capacity to do".

Organizational Capacity Findings

The second portion of the interviews conducted in this phase focused on the use of the framework. Participants were asked to identify the level at which their organization was located for each of the sub-dimensions. This proved to be a difficult exercise and highlighted the need for differentiation between some of the levels (as described in the previous section). The data analysis conducted focused on grouping the interview data by organization, while coding it by participant role (head of evaluation, senior evaluator, and evaluation user). A summary of the findings for each of the four organizations involved in the study is provided in Table 5 below and these are described in more detail in the pages that follow. Of particular note are the differences between the three individuals interviewed in terms of their assessment of their organization's

evaluation capacity; these differences may be due to the different knowledge levels of each individual based on their proximity to evaluation and their role within the organization. They could also be due to continued issues with the clarity of the framework and its interpretation by different individuals. This issue will be considered more thoroughly in the remainder of this chapter and in Chapter 7.

Table 6: Summary of Findings by Organization

Dimensions of EC	Org 1 (High)	Org 2 (High)	Org 3 (Int)	Org 4 (Low)
Human Resources	<p>Exemplary:</p> <ul style="list-style-type: none"> Experienced, highly skilled team and leadership Built-in career progression process Learning plans for all staff members 	<p>Exemplary:</p> <ul style="list-style-type: none"> Built-in career progression process Excellent PD opportunities Staff works in terms of expertise balance in terms of expertise Strong leadership 	<p>Intermediate:</p> <ul style="list-style-type: none"> Built-in career progression process PD opportunities depend on workload Highly specialized staff with good balance of skill sets Strong leadership 	<p>Developing:</p> <ul style="list-style-type: none"> Several vacant positions due to recent departures Mix of experience and project management skills within team Some PD opportunities for staff members Leader has strong evaluation and management background
Organizational Resources	<p>Intermediate:</p> <ul style="list-style-type: none"> Well resourced evaluation unit, expansive budget Data rich, but often low quality Many reorganizations but clear accountability lines Some issues with organizational supports 	<p>Intermediate:</p> <ul style="list-style-type: none"> Well resourced evaluation unit, additional funding available when needed No overall data collection system Stable governance structure, decentralized organization Good organizational supports, with some "irritants" 	<p>Intermediate:</p> <ul style="list-style-type: none"> Budget is sometimes shared with other business lines within the group Currently developing PM system but not implemented yet Good linkages to other corporate areas 	<p>Intermediate:</p> <ul style="list-style-type: none"> Stable funding source Data rich but difficult to access Clear organizational structure and accountability lines Organizational supports hinder timely production of evaluation work
Evaluation Planning and Activities	<p>Exemplary:</p> <ul style="list-style-type: none"> Extensive consultation for planning purposes Annual report on evaluation activities Good balance between in-house and use of consultants Location of evaluation unit allows for information sharing with other groups and good organizational linkages Extensive external supports; strong peer review system 	<p>Intermediate:</p> <ul style="list-style-type: none"> Extensive consultation for planning purposes Logic model for evaluation unit developed and use of report card Extensive contracting out of projects based on available human resources Staff encouraged to share info and lessons learned on projects No advisory or steering committees to maintain independence 	<p>Intermediate:</p> <ul style="list-style-type: none"> Five-year plan in place; some needs assessment done No systematic review of evaluation unit Exemplary model for use of consultants External networks could be improved Excellent access to senior management 	<p>Developing:</p> <ul style="list-style-type: none"> Evaluation plan based on some consultation with managers Consultants generally not used Some opportunities to share info among staff members Advisory committees and peer review process under development Good access to senior management
Evaluation Literacy	<p>Exemplary:</p> <ul style="list-style-type: none"> Involvement of key stakeholders through use of advisory committees Demand for evaluation from DG and ADM 	<p>Exemplary:</p> <ul style="list-style-type: none"> Involvement of stakeholders at key points in process Some confusion because of dual roles of PM facilitation and evaluation follow-ups 	<p>Intermediate:</p> <ul style="list-style-type: none"> Program managers involved at key points in process Evaluations thoroughly discussed at DG and ADM level 	<p>Developing:</p> <ul style="list-style-type: none"> Little awareness of evaluation outside senior management but building Good measurement of outputs, but difficulty in

<p>Organizational Decision-Making</p>	<ul style="list-style-type: none"> • Sign-off authority on RMAFs, with PM specialists within program areas <p>Intermediate:</p> <ul style="list-style-type: none"> • Evaluators often consulted on TB submissions and MC for pilot projects • Evaluation findings are used for decision-making when available • Some timeliness problems for use 	<ul style="list-style-type: none"> • Strong RMAF development process within program areas <p>Exemplary:</p> <ul style="list-style-type: none"> • Program staff require help with TB submissions • RMAFs used extensively • Evaluation info used for allocation and reallocation of funding 	<ul style="list-style-type: none"> • Evaluation unit often drives RMAFs <p>Intermediate:</p> <ul style="list-style-type: none"> • Use of evidence to support MC requirements • Increasing recognition of the contribution of evaluation to decision-making 	<p>measuring outcomes</p> <ul style="list-style-type: none"> • Just started doing RMAFs, evaluators take the lead <p>Intermediate:</p> <ul style="list-style-type: none"> • Growing demand for evaluation services from senior management
<p>Learning Benefits</p>	<p>Intermediate:</p> <ul style="list-style-type: none"> • Impact and level of implementation of evaluation recommendations not currently known • Lessons learned unit produces information packaged by theme (impact yet unknown) • Very transparent process in terms of publishing reports • Evaluations used to train new program staff 	<p>Intermediate:</p> <ul style="list-style-type: none"> • Some evidence of evaluation use, but only partial for now • No real evidence of process use as of yet 	<p>Intermediate:</p> <ul style="list-style-type: none"> • Evaluation findings sometimes used • No systematic follow-up on recommendations • Senior managers show some process use 	<p>Developing:</p> <ul style="list-style-type: none"> • Good implementation of evaluation recommendations • Will build on process use once advisory committees are up and running

Organization 1 – Exemplary Level

This service organization was identified as exemplary by the Phase 1 participants because of its investment in technical capacity and the rigour of its evaluation products. It has a large team of evaluators, divided up into a number of units organized by program area. This way, evaluators in each team become experts in a particular subject area and are able to gain a deep understanding of a small number of programs.

The three individuals interviewed in the organization were the Head of Evaluation (Director General level), the leader of one of the teams (Director level), and an evaluation user from a program area (Director level). All three participants commented on the organization's capacity for each of the dimensions included in the framework. Their comments are summarized in the following paragraphs.

Human Resources. The Head of Evaluation and team leader both felt that their organization showed exemplary capacity for all five sub-dimensions included in this section of the framework. The way in which each evaluation team is staffed allows for a range of skills and competencies to be used at various parts of the evaluation process. For instance, a senior evaluator leads a specific project and directs the activities of a number of more junior evaluators. These junior evaluators receive mentoring and coaching from the senior evaluator as they complete their assigned duties, and learn about evaluation in this manner. A career progression path is in place for certain positions, where it is possible for an evaluator to move up from one level to the other without the competition process used for promotion elsewhere in government: “everybody here is taken care of” (Evaluator 1). Promotions are instead based on attaining certain levels of a competency profile designed within the organization: “people are promoted according to knowledge” (Evaluator 1). All evaluation staff members have learning plans

tailored to their specific needs and are provided with internal and external training opportunities on a regular basis. In addition to this, individuals are encouraged to work with consultants in the field and can also move between teams for special assignments. Finally, the leadership of the evaluation unit was thought to be exemplary, with a Director General experienced in evaluation and competent directors to lead each team. From the perspective of an evaluation user, however, the picture is slightly different. Although this participant acknowledged the highly skilled workforce present in the evaluation unit, he also lamented the fact that evaluation products seem to take too long to be used by the organization's senior managers due to a mismatch between the informational needs of managers and the need to evaluate programs soundly and rigorously. There was also some concern regarding the level of program knowledge that evaluators have and the impact of this on the recommendations made in evaluation studies. To this participant, there is still room for improvement in terms of the mutual understanding between evaluators and program staff. Evaluators do not always take advantage of the existing knowledge on a particular program, whether "in people's heads or on paper".

Organizational Resources. The participants generally agreed on the sub-dimensions included here. They felt that the current budget situation of the evaluation unit is exemplary, with a stable, well-resourced budget obtained from programs (through their TB submissions) and from the Treasury Board directly. This represents a change from previous years, when the budget for the evaluation unit was entirely B-based (non-continuing funding) and there was ongoing concern about the long-term livelihood of the evaluation function within the organization. One area in which there is substantial room for improvement in this organization is with ongoing data collection. All three participants interviewed stated that although the organization collects large amounts of administrative data from its stakeholders, there are some quality issues and it is

difficult to integrate these data into evaluation studies and other performance reports. Some steps have already been made to remedy this situation, however: data assessments are now part of all evaluation studies, and a follow-up process on evaluation recommendations is currently being implemented in order to verify whether quality issues are resolved by the program areas. Finally, the organizational capacity in terms of its infrastructure was thought to be intermediate by all three individuals interviewed. A stable governance structure within the directorate has made it possible to draw clear accountability lines in most instances, although the organization has undergone several reorganizations in recent years. A special team within the evaluation unit responsible for contract management, budgeting, planning and HR makes it easier for evaluators to get their work done within established timeframes and “removes barriers to the effectiveness of the group” (Head of Evaluation 1). A communications team located within the evaluation unit also makes it possible to publish evaluation materials in a timely manner and establishes linkages with other communications teams located within program areas to coordinate the work that needs to be done.

Evaluation Planning and Activities. All three participants agreed that the evaluation planning process in this organization as well as the use of external supports made by evaluators reflect exemplary capacity, while the evaluators’ use of consultants, information sharing and organizational linkages reflect an intermediate level of capacity. Exemplary practices in evaluation planning include the establishment of a long-term plan based on consultations with all Directors General and an internal peer review process built into the evaluation plan to meet the organization’s high quality standards. An external peer review process is considered the cornerstone of this organization’s approach to evaluation and reflects exemplary practice in this area. This process involves consultation with academic experts on evaluation design and report

quality on an ad hoc basis, as well as more structured “methodological conferences” in which a number of experts are invited to a debate on a specific methodology. Consultants are typically used to conduct evaluation studies and to report on findings, while the study design is left to evaluation staff members. In the eyes of the participants, this allows the organization to capitalize on the expertise of its internal staff while maintaining its objectivity and neutrality. Some portions of the work can be done in-house, depending on the nature of the task to be completed and the workload of the evaluation team. Information sharing is thought to be excellent within each evaluation team; the Director responsible for each team keeps a close eye on the projects underway and shares progress with the rest of the management team during regularly scheduled meetings. A “Knowledge Management” group has also been established within the evaluation unit “to produce documentation on evaluation results for the consumption of various audiences” (Head of Evaluation 1). For example, this team can group the results of several evaluations by themes to inform policymaking. Finally, strong organizational linkages have been established by the evaluation unit through the use of evaluation steering committees that include program managers and other staff members. There is also constant contact between the evaluation unit and the organization’s senior management team, although the evaluation unit’s positioning within the organization (i.e., in the Strategic Policy Branch) may not provide it with the exposure it needs in terms of the broader organization.

Evaluation Literacy. All three participants felt that the organization has reached an exemplary level of capacity for this dimension. In terms of involvement in evaluation, the Directors General and Assistant Deputy Ministers are highly knowledgeable about evaluation and request the services of the evaluation unit on a continuing basis. The Director General of the evaluation unit sits on two policy coordination committees (one with DGs and another with

ADMs) and so has the ear of senior management on issues of importance to the organization. According to one of the participants, evaluators also provide advice and information to programs on specific issues not necessarily related to evaluation, such as program information contained within an evaluation report. They also respond to requests from programs about their Treasury Board submissions and other cabinet documents that have an evaluation or results component. The organization's strong results-management orientation is made evident by the presence of specialists at the program level who work on program indicators, logic models, and RMAFs. In fact, RMAFs are developed by program personnel, although the evaluation unit retains a sign-off authority to ensure high quality and feasibility.

Integration with Organizational Decision Making. An intermediate level of capacity was identified for this section of the framework, although some exemplary practices were also highlighted by participants. Evaluators are routinely consulted on Treasury Board submissions, although the focus of the consultation is usually on the estimated resources required to evaluate the initiative proposed in the submission. An exemplary capacity level would likely include a more in-depth consultation focusing on the expected results of the proposed initiative as well as a more elaborate description of potential evaluation issues. The decision support sub-dimension reflects exemplary capacity when evaluation findings are available at the appropriate time for decision making, which is not always the case. Generally, however, discussions on new programs or on program renewal automatically include a review of past evaluations, and the Deputy Minister usually insists that evaluation findings be a part of the decision making process on all Grants and Contributions programs.

Learning Benefits. Participants agreed that the organization has attained an intermediate level of capacity for this dimension. As mentioned previously, evaluation findings and

recommendations are used in decision making whenever they are available. The evaluation unit is developing a standardized follow-up process to assess the implementation of its recommendations, but no information is currently available that demonstrates to what extent this happens after an evaluation report is released. The Knowledge Management group provides access to evaluation products and makes their assimilation into policy easier, but again, it is difficult to judge to what extent this occurs in the different program areas. Evaluation reports are also often used as training materials for new program staff members, because they contain streamlined information about program design and expected results. Finally, there is currently little evidence of process use within the organization, although one of the participants stated that evaluation reports “forces us to think about whether this (program) was a good idea in the first place” (Decision-Maker 1).

Observations and Summary of Findings for Organization 1. All three participants generally agreed on the level of capacity of the organization on the various sub-dimensions included in the framework. The evaluation user brought a slightly different perspective to the organizational assessment by acting more as a “critical friend” and pointing out some of the evaluation unit’s flaws. The fieldnotes taken during the three interviews conducted in this organization point to a certain level of disconnect between the evaluators and evaluation user in terms of evaluation use: while the two evaluators spoke highly of the extent to which evaluation recommendations are implemented in the organization, the evaluation user expressed some concern about the potential weight of evaluation recommendations without stating this in so many words. This participant spoke often of the involvement of evaluators in policymaking recommended in the upcoming evaluation policy and suggested that this may be an area in which evaluators would be overstepping their bounds. He used words such as “painting government

into a corner” to reflect the fact that evaluation should be only one of many decision-making instruments to be used by senior decision-makers, and that other factors beyond those included in evaluation recommendations need to be considered before some policy decisions are made. He clearly voiced his concern about the amount of power evaluation recommendations could have in the future, and advocated the production of softer recommendations that did not cause the department to commit to a specific set of actions. His views on this issue are based on his perceptions that evaluators don’t have the full picture because of their external stance towards the program. This discomfort with the types of recommendations produced in evaluation may suggest some power imbalances in the policymaking process within the department; this issue will be explored to a greater extent in the next chapter.

Aside from these organizational findings, the framework seemed to pose no significant problems in terms of clarity or interpretation for the three participants; however, the high level of capacity demonstrated by this organization may have made its use easier. The findings support the predictions made by the participants in the first phase of the study, who felt that this organization showed an exemplary level of capacity. This was certainly found to be true for the “capacity to do” dimensions, and the “capacity to use” dimensions reflect the organization’s progression from intermediate to exemplary levels of capacity.

Organization 2 – Exemplary Level

This regulatory organization has been identified by the Phase 1 participants as having exemplary capacity. This rating was given based on the evaluation unit’s continued efforts in building organizational capacity by reaching out to program areas and by diversifying the products offered to its clients. This organization has received awards for these efforts and is often held up as an example to follow for other government departments and agencies. Three

individuals were interviewed in this organization as part of the third phase of the study: the Head of Evaluation (Director level), a senior evaluator, and a program specialist responsible for evaluation. This third participant works directly within a program group and manages evaluation-related projects. However, this person's definition of evaluation is somewhat different than the generally accepted notions of evaluation held across government. He is responsible for the development of "evaluation protocols", a set of guidelines used in the assessment of applications made to the department for funding. In other words, he provides assistance towards program delivery and not program evaluation. Nonetheless, he is the primary contact for evaluation within his group and his stance on evaluation reveals some discomfort with the more traditional model of a central evaluation function. According to this individual, the traditional model does not make good use of the program knowledge that exists within program areas; rather, evaluators must learn what they can on a particular program while they work on its evaluation. In the view of this participant, this leads to poor recommendations that are of little use to program staff, both because of the delays common to most evaluation studies and also because the recommendations often reveal a lack of knowledge about important program details. In his opinion, evaluation reports only provide information that is already known to program staff and therefore are only useful as far as they help programs meet Treasury Board requirements. The participant chose not to use the framework as part of the evaluation capacity assessment and instead commented on specific sub-dimensions of the framework. For instance, he felt that the existence of a career progression path for evaluators did not reflect exemplary capacity because in his view, staff members at any level benefit from moving to a different job after 12 to 18 months in a given position. This is somewhat contradictory to the development of in-depth program knowledge that he feels is critical to quality evaluation products. In the end, his

comments may have reflected a lack of understanding of governmental evaluation principles and practices. It is unclear why this individual was recommended as a potential participant; he had recently worked with evaluators responsible for a review of the program in which he is involved, and perhaps this was the reason behind his nomination by the Head of Evaluation. The paragraphs that follow on the specific dimensions therefore only reflect the data collected from the head of evaluation and the senior evaluator who participated in the study.

Human Resources. The organization has spent substantial time and effort in building up its human resources capacity in terms of staffing and professional development. Therefore, participants found that the organization is well described by the intermediate or exemplary levels of all five sub-dimensions included in this section of the framework. In terms of staffing, the organization has reached a good balance between junior and senior evaluators and has implemented a career progression process for junior evaluation professionals, in order to provide them with the opportunity to move towards higher classification levels without a formal competition process. In terms of technical skills, most evaluators are generalists rather than specialists in particular methods, and so the participants felt that further training could benefit the unit. The communications skills, professional development and leadership were all thought to reflect an exemplary level of capacity.

Organizational Resources. An intermediate level of capacity was thought to best reflect the current state of this organization in terms of the availability of resources for evaluation. The budget allocated to evaluation was thought to be appropriate given the needs outlined in the plan, with additional funding provided when unexpected expenses arise. Programs also request and set aside funding for evaluation projects. Ongoing data collection varies from one program to the next, but most programs now have performance measurement frameworks and are in the early

stages of implementing them. There is no organization-wide system to which programs contribute data, most likely because of the size of the organization and its decentralized approach to management. Therefore, the capacity level of the organization on this sub-dimension was thought to be low by both participants. The organizational infrastructure reflects an intermediate capacity level, with a stable governance structure and organizational supports that provide an adequate service, although evaluators are sometimes inconvenienced because of time delays: “we have good supports, just some irritants like procurement policies and HR processes” (Head of Evaluation 2). The department has no evaluation policy of its own, and there is currently no intention of developing one. The feeling within the evaluation unit is that the Treasury Board Evaluation Policy “is enough if you need to swing a mallet at somebody” (Head of Evaluation 2).

Evaluation Planning and Activities. The participants identified exemplary practices in terms of the evaluation planning process. In addition to in-depth consultation with senior managers and the existence of a multi-year plan, the evaluation unit itself engages in evaluating its own impacts on the organization by using a logic model and issuing a report card for the unit every year. The other sub-dimensions included in this section of the framework were thought to reflect intermediate capacity levels. For instance, although the goal is to only use consultants when absolutely necessary, the current state of affairs requires the evaluators to “contract out more than we would like” (Head of Evaluation 2). Staff members are encouraged to share evaluation findings and lessons learned during projects to others during formal meetings, but beyond that, no other processes are in place to facilitate information sharing. Although the evaluators make frequent use of external supports such as professional associations, which reflects exemplary capacity, they do not generally liaise with stakeholders and do not keep

themselves informed of new developments in other organizations. In this sense, their focus remains inwards rather than turned towards the external environment of the organization. Finally, the organizational linkages established within the department were thought to reflect intermediate capacity, with fairly good access to senior managers and good contact with program managers. However, because of the decentralized approach used to structure the organization, there is little proximity to other groups such as policy or strategic planning.

Evaluation Literacy. The involvement in evaluation of program staff and senior managers was thought to reflect an exemplary capacity level. Those involved with evaluation “figure out what’s going on” (Head of Evaluation 2) and participate in the evaluation process to some extent. However, within the broader organization, people are not necessarily familiar with evaluation and some marketing is still needed to increase the visibility of the unit: “there is still some confusion due to our dual role in facilitating performance measurement and conducting evaluation follow-ups” (Evaluator 2). In terms of results-management orientation, the organization shows intermediate capacity on many fronts, with a strong results-based management culture throughout most of the organization and logic models developed for almost all programs. Program areas have the responsibility to develop and implement their own RMAFs, with occasional assistance from evaluators on technical matters.

Integration with Organizational Decision-Making. Management processes within the organization take advantage of the expertise of evaluators when developing Treasury Board submissions. In this area, the organization shows exemplary capacity: the evaluation unit is also consulted on the development of Memoranda to Cabinet and other performance documents such as the Program Activity Architecture and program RMAFs. The same level of capacity was found to be true for the decision support sub-dimension, especially within the context of a special

review done by Treasury Board in which “poor performers” were identified based on evaluation information.

Learning Benefits. The two different types of use were identified as reflecting an intermediate level of capacity for this organization. Participants were unsure to what extent evaluation findings and reports have an impact on stakeholders’ understanding and attitudes about programs, although they could point easily to examples of evaluation recommendations being implemented to improve a program. Some evidence is also available on process use; however, participants rated it as “not strong enough yet” to assess the organization’s capacity level as exemplary for this sub-dimension.

Observations and Summary of Findings for Organization 2

The use of the framework in this organization was done in a very methodical and meticulous manner, with two of the three participants providing detailed comments about each element (bullet) included under each of the sub-dimensions. In this way, the organizational assessment that was done resulted in a more nebulous picture of the organization’s evaluation capacity, since many of these elements were not all found to reflect the same capacity level. However, as in other cases, the general impression of intermediate to exemplary capacity was maintained throughout the assessment and thus allowed the participants to make broad statements about their organization’s evaluation capacity. The most interesting aspect of this organizational assessment was undoubtedly the completely different stance taken by the evaluation user, who refused to acknowledge the traditional model of the central evaluation function that is reflected in the framework, opting instead to discuss an alternative model in which evaluators would be positioned within the program groups directly. The pros and cons of this approach were discussed during the interview, and at the end of the discussion, the

participant stated that he understood the researcher's point of view and approach, but felt that a change towards the alternative model would be beneficial to departments across government. As this is somewhat beyond the scope of the study, this will not be considered in the thesis, but raises interesting opportunities for future research on evaluation capacity. An interesting observation included in the fieldnotes for this interview is that the actual conversation did not take place in the evaluation user's office, as for the others. This person asked to meet the interviewer at his office, but then decided that a better place for the interview would be a public seating area in the lobby of the building. This might suggest a certain level of discomfort in terms of showing participation in the research project to his program colleagues. Throughout the interview, this participant seemed nervous and uncomfortable in his criticism of the central evaluation function, even though we were seated in a different building than the one housing the departmental evaluators.

Organization 3 – Intermediate Level

The individuals involved in this science-based agency include the Head of Evaluation (Director level), a senior evaluator and a decision-maker at the Vice-President level. The organizational assessment made using the framework was done mainly by the Head of Evaluation as well as the senior manager, whereas the feedback obtained from the senior evaluator focused almost exclusively on the contents of the framework itself.

Generally, the comments made by the Head of Evaluation and decision-maker were similar in nature. Both participants were able to recognize the strengths of the organization on certain dimensions and identify certain areas for improvement on others. As could be expected, the information provided by the Head of Evaluation featured operational details and focused almost exclusively on the first three dimensions (capacity to do). In contrast to this, the

information provided by the decision-maker focused more heavily on evaluation use and provided a complementary perspective on some operational issues.

This organization had been identified as being a strong intermediate by the individuals interviewed in the first phase of the study. These participants recognized recent efforts made by the Head of Evaluation to build the organization's evaluation capacity on several fronts, such as staffing, professional development, ongoing data collection, and management processes. These impressions were confirmed by the third phase participants, who were able to provide additional details regarding these continuing efforts. These details are summarized by dimension in the following paragraphs in an effort to demonstrate how the framework was used in this organization.

Human Resources. Overall, this dimension was rated as "high-intermediate" by both the Head of Evaluation and the decision-maker. The exemplary level of all five sub-dimensions was thought too high at the current time, although various initiatives have already been put in place to reach it in the future. For instance, although not all positions are filled, new positions are currently being created in order to recruit individuals with specific backgrounds and experiences. New staff members have been recruited for their skills in certain areas, since most of the evaluation work is done internally. Career progression has been standardized for evaluators, with evolving responsibilities built into the classification and performance appraisal process. Both the Head of Evaluation and the senior manager felt that the senior evaluators demonstrated skill levels that were well-described by the exemplary level, both in terms of evaluation logic and communications skills. One area identified as weaker by the Head of Evaluation was the Professional Development sub-dimension. Although some efforts have been made to address this

(e.g., brown bag sessions, dry runs of paper presentations), this participant felt that more could be done, such as peer review between senior evaluators.

Organizational Resources. Participants felt that the organization was well described by the intermediate levels of the sub-dimensions included in this section of the framework. For example, although the evaluation unit receives continuing A-base funding, the specific amounts included in the budget sometimes vary. This is due to the fact that the evaluation unit is situated within a larger group of three different business lines. The budget is allocated to this larger group, and some funds occasionally get reallocated to other business lines depending on need. There is some overlap between the responsibilities of the different business lines (e.g., the Departmental Performance Report falls under the responsibility of the Performance Measurement Group, yet is included under the budget of the evaluation unit), and staff members sometimes move between business lines. This makes it more difficult to allocate budget dollars in a stringent way and therefore requires more flexibility in how money is spent. Ongoing data collection was also situated at the intermediate level by participants: although a performance measurement system exists, it is still somewhat unrefined and requires a more in-depth implementation in order to provide useful data to evaluators. For instance, the information provided annually by various program groups is static and does not reflect changes in direction or other important decisions. A corporate scorecard that includes targets as well as roles and responsibilities has recently been introduced to senior management and shows some promise in terms of providing an impetus to programs for better ongoing data collection. Finally, the organizational infrastructure provides a sound accountability path and a clear governance structure. There has been some discussion regarding the development of an organizational evaluation policy but no work has been done as of yet to move this issue forward. Both the Head

of Evaluation and senior manager felt that other organizational groups, such as procurement, HR and communications, provide a good service and pose no real problem for evaluators.

Evaluation Planning and Activities. Some of the sub-dimensions included in this section of the framework were identified as “exemplary” for this organization, while the others reflect the organization’s capacity at the intermediate level. The organization was found to have exemplary capacity in the “Use of Consultants” and “Organizational Linkages” sub-dimensions, because of the highly skilled evaluation staff on the first and because of the ready access that evaluators have to the Vice-Presidents and President of the organization on the second. The evaluation unit is located close to the policy, performance measurement and strategic planning groups, and evaluation is now built into corporate planning processes, which speaks to its relevance to the main corporate agenda. The organization demonstrates intermediate levels of capacity in terms of its evaluation plan, since risk assessment poses some difficulty to the evaluators (due to stringent Treasury Board requirements), and also because no systematic review of the evaluation unit is in place. The information sharing done within the evaluation unit as well as its use of external supports were also found to be at an intermediate level: although some efforts have been made to improve these two aspects, some work remains to be done in order to move to an exemplary level.

Evaluation Literacy. The organization was found to demonstrate an intermediate level of capacity on both sub-dimensions in this section of the framework. Although some results-management orientation can be identified at the upper echelons of the organization, this has not yet been transmitted down to the other levels and therefore, results in somewhat limited awareness and understanding of results-based management principles and practices throughout the organization. Program managers are involved at key points of the evaluation process,

however, and they have the opportunity to share their views on evaluation findings and recommendations in their management responses. This organization shows high evaluation literacy at the senior management level; evaluations are thoroughly discussed at the Director General and Vice-President levels because, according to one participant, “they’re on the hook at senior management committee to defend their responses to the recommendations” (Decision-maker 3). In fact, the senior management committee (the organization’s executive committee) is the evaluation unit’s primary steering committee.

Integration to Organizational Decision-Making. Only the decision-maker was able to comment on this section of the framework due to his role in the organization. Without specifying a particular level on either sub-dimension, this participant provided some information about the decision-making process in the organization and the place of evaluation within this process. New initiatives have clearly linked evaluation to the decision-making process; for instance, a new set of programs have integrated a formative evaluation mechanism to their lifecycle and the information produced through these evaluations feeds directly into the requirements set out by the Memorandum to Cabinet (MC). In other words, the evaluation results are used to allocate resources to these programs and provide information used for business planning. As the organization matures, there is a feeling that evaluation is increasingly recognized for its contributions to decision making. A systematic follow-up to the implementation of evaluation recommendations is also planned for the near future.

Learning Benefits. Once again, the participants felt that an intermediate level of capacity has been achieved for this dimension. As just mentioned, no systematic follow-ups on evaluation recommendations are currently done; therefore, only anecdotal information is available regarding the use of evaluation findings. Evidence of process use is visible within the senior management

ranks more than at the program level. Senior executives need to be results-based, according to the participants, and therefore, need to ask “how are we going to measure this” when new initiatives are proposed.

Observations and Summary of Findings for Organization 3. The framework was used consistently from one participant to the next within this organization. The general consensus that the organization has an intermediate-to-exemplary capacity level reveals that the descriptions of the sub-dimensions were interpreted in a similar fashion across individuals and provided sufficient differentiation between the levels to be considered useful by users. It also supports the original assessment made by the participants of the first phase of the study. It should also be noted that the organization as a whole did not have a uniform level of capacity for all of the dimensions.

Organization 4 – Low Level

This organization has an operational mandate within the federal government and has therefore only recently been required to evaluate its programs. Because it is still in the early stages of developing a central evaluation function, its capacity level was found to be low by the Phase 1 participants, who cited a high turnover rate and a low skill level as the main reasons behind their assessment. Three participants were contacted as part of the third phase of the study: the Head of Evaluation (Director level), a senior evaluator, and a corporate-area (i.e., non-program) evaluation user (Director level). Although the Head of Evaluation and senior evaluator readily agreed to participate in the study, the evaluation user was more difficult to reach and subsequently chose not to participate. Further attempts at identifying other potential participants were unsuccessful. The results of the study for this organization therefore reflect the opinions of the two evaluation professionals interviewed, with no substantive input from evaluation users.

This is an important issue both in terms of assessing the quality of the data obtained (i.e., the ratings on the sub-dimension were made by people closely involved with evaluation rather than more impartial observers) and in terms of revealing the organization's true capacity (i.e., the fact that no one from programs could be identified as a potential participant speaks volumes about the organization's awareness and interest in evaluation).

Human Resources. Both participants had some difficulty with this sub-dimension due to differentiation issues between the intermediate and low capacity levels. They felt that the low capacity level was too low for their organizational situation, and that the intermediate was too high to properly describe its status. This issue has been resolved in the final version of the framework with the addition of the "developing" capacity level, largely based on these participants' comments. An important contributor to the relatively low capacity level of this organization in terms of its human resources is the high turnover rate experienced in the organization in the past few years. The evaluator classification group was changed from the "Administrative Services" group to the "Economics and Social Sciences" group recently. This has in effect changed the job description of evaluators and increased the technical knowledge and experience required to work as an evaluator. This has resulted in five departures last year alone and has left the organization with more positions than it can afford to fill. The evaluators still working in this organization have varied levels of experience and technical skills, and the professional development in which they take part meets the intermediate description of the framework. One asset of this organization is that the leader of the evaluation unit has a strong management background and has been working on developing strong relationships with the organization's senior managers.

Organizational Resources. Budget resources available for evaluation were felt to represent an intermediate to exemplary level because of the stability of the funding that is made available to the evaluation unit. Ongoing data collection was also in the intermediate range, with a significant amount of available administrative data for use in evaluations. These data are used in quarterly reports from programs presented to the management board of the organization, but are not easily translated into concrete evaluation data. The large size of the organization has resulted in a stable organizational structure with clear accountability lines. The organization has recently implemented its own evaluation policy, although more time is needed before its impact can be properly assessed. Finally, some other corporate organization supports, such as HR, procurement and communications “tend to hinder the timely production of evaluation work” (Evaluator 4).

Evaluation Planning and Activities. The organization was found to fit within the intermediate descriptions of most sub-dimensions included in this section of the framework. The evaluation planning process is fairly thorough and involves consultation with some managers. There is a risk assessment process in place to assist in priority setting within the plan but no ongoing intelligence gathering on new developments within the organization. Participants felt that this could be done to move the organization to an exemplary capacity level. Most of the evaluation work is conducted in-house, with very little use of external consultants. The participants felt that this was reflective of exemplary capacity, but given their earlier comments on the current level of evaluation expertise within the unit, a more balanced distribution of internal and external resources to evaluation projects may be appropriate and desirable in this case. Information sharing was also thought to reflect an intermediate capacity level, with ongoing opportunities to share information among staff members and efforts to establish a peer

review process as well as advisory evaluation committees within the organization. Again, the “intermediate” rating may be more reflective of the missing level between the low and intermediate levels than the appropriateness of the description included in the intermediate level for this sub-dimension. The use of external supports was thought to be situated between the low and intermediate levels, given the administrative nature of the organization and its inward looking focus. Finally, the organizational linkages made between evaluation and other corporate units was thought to be close to exemplary, with direct reporting on evaluation to the organization’s executive committee and ties to the decision-making process on new initiatives.

Evaluation Literacy. As evaluation is fairly new to this organization, both participants felt that the low capacity level best reflected both sub-dimensions included in this section of the framework. They felt that there was very little awareness of evaluation within the broader organization, even though evaluation is directed towards senior managers on a more consistent basis than it has ever been. One participant felt that some in-roads have been made in the past five years, however, in that most organizational members can now distinguish between program evaluation and audit. The results-management orientation of the organization also leaves to be desired in the view of the participants, especially in terms of measuring outcomes. The organization has a long history of measuring outputs, but has difficulty thinking in terms of outcomes. The evaluators have the responsibility for the development of RMAFs, which are completely new to the organization.

Integration with Decision-Making. A growing demand for evaluation services speaks to the intermediate level reached within this organization in terms of its management processes. There is still little evidence that evaluation findings are used for decision support, however, but participants were hopeful that this is slowly changing.

Learning Benefits. The recommendations made in evaluations are generally well implemented within this organization. This speaks to intermediate-level conceptual and instrumental uses of evaluation, which are often precursors to more sophisticated process use. On this front, the participants felt that organizational capacity was quite low for the time being, since program staff members are not generally involved in evaluation projects. This may change in the future, once evaluation advisory committees are created.

Observations and Summary of Findings for Organization 4. The lack of an evaluation user for this organization has resulted in findings that reflect only the opinions of those closely involved with evaluation. Although these individuals provided their honest opinion on their organization's various levels of capacity on each of the sub-dimensions, this analysis would have been made far richer with the inclusion of an external point of view. However, the fact that no evaluation user could be found to participate may demonstrate a lower capacity level than that described by the evaluation professionals, as mentioned in the introductory paragraph to this section. Overall, the participants for this organization felt that an additional level would have been necessary to properly describe their organization's evaluation capacity, since "low capacity" did not adequately reflect the efforts made by the organization to increase its capacity, and "intermediate capacity" was often found to describe a higher level of capacity than that of the organization. An additional level named "developing capacity" has been included in the final version of the framework to remedy this situation. This finding is in line with some of the fieldnotes taken during these two interviews, which highlight the positive nature of the comments made by the Head of Evaluation and the less positive comments made by the Senior Evaluator. These comments could be due to each participant's tenure within the evaluation unit (i.e., the Head of Evaluation is relatively new while the Senior Evaluator has been in this unit for

a significant number of years) or could also be due to each person's role within the unit (i.e., Managers are often thought of as cheerleaders for their teams, while team members do not have that type of role and can perhaps provide a more objective assessment). In any case, the fact that both participants, for the most part, cast their organizational evaluation capacity in a fairly positive light (in terms of stating that there should be a "developing capacity" level because their organization does not fit within either the "low" or "intermediate" levels) may suggest a deeper concern with optics and how their organization may be perceived in the larger evaluation community. This issue will be revisited in the next chapter.

Observations on Organizational Capacity Findings

Although the research questions that guide this study do not focus specifically on the differences in evaluation capacity that might exist between various federal government organizations, a comparative analysis undertaken as part of the framework validation exercise has yielded some interesting findings. First, it should be noted that the holistic assessments made for each organization by the participants using the framework resemble those made by the evaluation experts consulted as part of the organization selection process. That is to say that the impressions of the evaluation experts in terms of identifying high, intermediate and low capacity organizations seem to be consistent with the capacity levels selected for most of the six dimensions of capacity by individuals working in those organizations. This observation provides preliminary evidence of the validity of the framework, since the correspondence between the qualitative assessments provided by the evaluation experts and the departmental staff suggests that the framework is helpful in drawing an accurate picture of an organization's evaluation capacity. An interesting observation is that the general assessments of capacity provided by evaluation experts match exactly those obtained by using the framework for the Human

Resources and Evaluation Literacy dimensions. This suggests that the general assessment made by the four experts was mainly based on the elements included under these two dimensions. A further examination of the Phase 1 and 2 data revealed that this was indeed the case.

Second, as mentioned previously, the use of the framework by organizational members has revealed that no one organization has a consistent capacity level from one dimension to the next, and indeed, even between sub-dimensions. This shows that the framework highlights some of the capacity building efforts already undertaken and those areas that require more work in order to reach a higher capacity level. Therefore, the framework appears to be a suitable growth profile for evaluation capacity.

Finally, the fact that the four participating organizations have the same capacity level for some dimensions (e.g., Organizational Resources) suggests that some elements of evaluation capacity depend on the type of organization and their general structure. For instance, the budget and organizational structure elements of Organizational Resources reveal a similar situation in all four organizations. This is likely attributable to the fact that they are all federal government organizations and share similar budgeting processes and accountability structures.

Chapter 7 – Discussion

This chapter provides a summary of the key findings of the study and situates these within the broader literature on evaluation capacity building. More specifically, the three research questions guiding the study will be revisited and the implications of the findings on the conceptual framework outlined in Chapter 2 will be presented. Further implications for research and practice will also be presented at the end of this chapter.

Motivation for the Study

The question of evaluation capacity in governmental organizations has been studied empirically in only a limited fashion. The actual dimensions of evaluation capacity, or what evaluation capacity might look like in an organization, have not previously been identified in the literature. Most research articles on this topic focus instead on the steps taken to build evaluation capacity, and stop short of describing the various facets of evaluation capacity itself. Developing such a description was the primary motivation for this study. It can be argued that without a clear picture of what exemplary evaluation capacity looks like, any efforts made by organizations to increase their capacity are incomplete. A clear goal can help organizations determine what efforts need to be made, and the direction that these efforts must take.

Overview of Study

This study sought to answer three research questions: What are the essential dimensions of evaluation capacity in Canadian federal government organizations? How are minimal and exemplary performance on each of these dimensions characterized? What are the manageable steps required to move from minimal to exemplary performance?

A three-phased approach was used to produce answers to these questions. First, an in-depth literature review was conducted in order to identify some of the potential dimensions of

evaluation capacity as described in anecdotal reports of ECB. These data were supplemented with key informant interviews conducted with four evaluation experts with experience in conducting and managing evaluation projects on behalf of the Government of Canada, and who are also knowledgeable about the capacity level of the various departments and agencies that make up the federal government. These individuals were asked to identify some of the dimensions of evaluation capacity. They were also asked to provide examples of government organizations with exemplary, intermediate and low or typical capacity. The literature review and interview data were combined in a first analysis, which generated a draft version of a framework of evaluation capacity.

The second phase of the study consisted of another round of interviews with the same four participants as those included in the first round. The purpose of this round was to obtain feedback from participants on the draft framework. Participants commented both on the wording used in the framework as well as on its more conceptual components. Further revisions were made to the framework following this phase.

The third and final phase of the study focused on validating the framework in four federal government departments and agencies. Two of these organizations had been identified by Phase 1 and 2 participants as having exemplary capacity, one was identified as having intermediate capacity, and one was identified as having low capacity. Three individuals in each of the four organizations were interviewed (with the exception of one organization, in which only two individuals were interviewed because of administrative issues). In each case, an attempt was made to interview the Head of Evaluation, a senior evaluator, and an evaluation user to benefit from the different perspective that each individual might bring to the study. The participants were asked to comment on the clarity of the wording used in the framework, the differentiation

between the capacity levels outlined in the framework, and the actual contents of the framework (i.e., dimensions and sub-dimensions included in the framework). They were also asked to situate their organization within the various levels of evaluation capacity outlined for each sub-dimension of the framework. These data were analyzed and a final version of the framework was produced, in addition to a description of the evaluation capacity in each of the four organizations.

Summary of Findings by Research Question

The final version of the framework, presented in Chapter 6 (Table 4), provides a summary of the key findings of this study. A detailed description of these findings, organized by research question, is presented in the following paragraphs.

Dimensions of Evaluation Capacity in Government Organizations

The first research question guiding this study was: *What are the essential dimensions of evaluation capacity in Canadian federal government organizations?* Six main dimensions were identified through the three data collection phases of the study and divided into two broad categories, “capacity to do” evaluation and “capacity to use” evaluation. Most participants focused on the “capacity to do” category, likely because the dimensions included in this category are easier to control and speak to the more operational facets of evaluation.

The first dimension, Human Resources, refers to the composition of the evaluation unit itself and is divided into five sub-dimensions. The first sub-dimension, staffing, refers to the balance of evaluation positions within the organization and whether these are sufficient to manage the workload identified in the evaluation plan. It also includes career progression for evaluators, which deals with employee retention and succession planning, two issues crucial to capacity building and maintenance. The second and third sub-dimensions focus on the technical and interpersonal skills required by evaluators. Skills related to the identification of evaluation

issues, the use of appropriate data collection methods, the generation of evidence-based recommendations and project management are part of the technical abilities required of evaluators. Softer skills such as building client trust, communicating evaluation messages in a clear and transparent way, and meeting program stakeholders' informational needs are part of the communications and interpersonal skills used by evaluators. The fourth sub-dimension refers to professional development and includes elements related to both internal and external professional development activities, as well as the development of learning plans for evaluation staff members and ongoing assessments of the skill set that exists within the evaluation unit. Finally, the fifth sub-dimension included under Human Resources refers to the quality of the leadership within the evaluation unit. Good leaders should have evaluation and management experience, be able to translate the information needs of senior managers into concrete project plans and act as mentors or coaches for team members.

Participants focused heavily on the HR dimension during the interviews, especially those directly involved with evaluation. In fact, the overall assessment of the organizations provided in the first phase of the study appears to be closely linked to the later capacity assessment made of this dimension by organizational members. Therefore, it could be argued that the impression of evaluation capacity is mainly based on the Human Resources dimension rather than on a more balanced perspective that includes all six dimensions.

The second dimension outlined in the framework is named Organizational Resources and offers a complementary perspective to the first dimension. Three sub-dimensions are included in this section of the framework: Budget, Ongoing Data Collection, and Organizational Infrastructure. The budget sub-dimension refers to the stability of the evaluation budget and whether it provides sufficient funding to complete the activities outlined in the evaluation plan.

Ongoing data collection speaks to the performance measurement systems that are in place within the organization and that produce information that is fed into evaluation studies. The organizational infrastructure refers to the stability of the governance structure, the existence of organizational evaluation policies and the organizational supports that help or hinder the work of evaluators, such as procurement services.

The third dimension focuses on the activities undertaken by evaluators as part of their regular duties. The development of an evaluation plan is key amongst the sub-dimensions that make up this section of the framework. This sub-dimension is characterized by the development of an evaluation plan in consultation with other stakeholders, the inclusion of a risk assessment process in the identification of evaluation priorities, ongoing intelligence gathering, and a systematic review of the evaluation unit itself. Evaluators in most departments make some use of consultants to provide assistance with various matters. This was therefore included as a sub-dimension, since it is a good indicator of the organization's evaluation capacity. Information sharing within the unit was also included here, since evaluation staff members spend a considerable amount of time sharing information related to their progress on certain files or on general project management issues with their colleagues. Evaluators in some organizations also spend time establishing linkages with external supports such as professional associations, program stakeholders and other organizations likely to provide assistance on various matters, such as the Treasury Board Secretariat. In the same sense, evaluation staff members may also establish linkages within their own organizations through formal or informal ties in order to remain informed on policy decisions likely to affect their work and to better share the results of the evaluations conducted by members of the unit.

The fourth dimension is the first included under the “capacity to use” evaluation category and reveals a less operational perspective than the first three dimensions of the framework. This dimension focuses on evaluation literacy within the organization and is divided into two sub-dimensions: involvement in evaluation and results-management orientation. Involvement in evaluation refers to the participation of program staff and other stakeholders in the evaluation process. Participatory evaluation theory holds that the greater the involvement of stakeholders in all phases of an evaluation, the greater the instrumental, conceptual and process use of evaluation. Therefore, in order to build evaluation capacity, organizations must pay attention to the involvement of staff members in the evaluation process. Results-management orientation refers to the larger organizational culture and the messages that are brought forward by senior managers. A results-management orientation can be manifested through the development of results chains for programs and the implementation of performance management frameworks such as RMAFs.

The fifth dimension follows in the footsteps of the previous one and focuses on the integration of evaluation information with organizational decision-making processes. At the outset, the management processes of the organization such as the development of Memoranda to Cabinet and Treasury Board submissions should include evaluation considerations in order to ensure that sufficient resources are provided for the eventual evaluation of new initiatives. At the final stage of the evaluation process, the findings and recommendations made in an evaluation study should be clearly linked to budget allocation and other high-level organizational and policy decisions. An organization with exemplary capacity searches out evaluation information in its decision-making process and relies on this information on an ongoing basis.

Finally, the sixth dimension, learning benefits, provides information on the types of uses that can be made of evaluation information within an organization. At a more operational level, the evaluation findings can be used as a basis for action and change through the implementation of evaluation recommendations (instrumental use). The evaluation findings can also have an impact on the understanding and attitudes of stakeholders about a program by clarifying certain operational aspects or by highlighting specific program results (conceptual use). At a broader level, the participation of organizational members in the evaluation process can sometimes result in behavioural or cognitive changes within these individuals based on their exposure to evaluation (process use).

Capacity Levels – Minimal to Exemplary Performance

The second research question guiding this study was: *How are minimal and exemplary performance on each of these dimensions characterized?* The approach taken to develop the various capacity levels for all six dimensions and their sub-dimensions was based on Leithwood and Montgomery's (1987) Innovation Profiles method, in which a low capacity level and an exemplary capacity level are identified for each sub-dimension, and the manageable steps required to move from one to the other are then developed accordingly, to answer the third research question, *What are the manageable steps required to move from minimal to exemplary performance?* The first draft of the framework (Chapter 4, Table 2) included a variable number of levels for the different sub-dimensions, with some having three capacity levels and some having four. The levels of this early draft were named "Typical or Developing Capacity", "Intermediate Capacity", and "Exemplary Capacity".

The second draft of the framework (Chapter 5, Table 3) features only three levels throughout the various sub-dimensions. These levels are named "Low or Developing Capacity",

“Intermediate Capacity”, and “Exemplary Capacity”. The reduction of the number of levels for some sub-dimensions is based on the inclusion of additional information within each level, which made the move between levels more manageable in the eyes of participants. Some participants felt that a uniform structure (i.e., same number of levels throughout the framework) made it easier to understand and use the framework. The change in the name of the first level reveals the participants’ unease with the use of the word “typical”. Because this term was used to essentially describe what can be thought of as low capacity, participants worried that “typical” legitimizes low capacity and that the framework would lose some of its explanatory power if this term was left in the framework. The participants recommended the use of the word “low” to properly describe the state of minimal capacity, and this change was made based on this advice.

The final version of the framework (Chapter 6, Table 4) is based on the validation exercise conducted in four different government organizations. Here again, participants felt that a uniform number of levels made the framework easier to use while providing sufficient information. However, a fourth level was added to the framework after several participants experienced some difficulty in differentiating between the low and intermediate capacity levels. Participants felt that the “Low and Developing Capacity” level actually represented two levels of capacity, one in which no effort has yet been made in terms of evaluation capacity building, and one in which some effort, albeit limited, has been made to develop the organization’s evaluation capacity. The terminology employed in the final version of the framework features “Low Capacity”, “Developing Capacity”, “Intermediate Capacity”, and “Exemplary Capacity”.

The specific elements included in each capacity level also varied significantly over the course of the development of the framework. The first draft of the framework included only one element, or bullet, in each capacity level, while the final version includes several elements for

each level. Elements were added as necessary to increase the clarity of the description and to properly differentiate between levels. It is highly probable that not all elements within one level describe an organization perfectly. However, even though the use of levels implies a certain amount of measurement, and therefore, judgment on organizational or individual performance, the purpose of the framework is not to measure but to describe organizational evaluation capacity. The intent of the framework is to provide organizations with information that can be used to identify the particular elements that they need to improve upon in order to fully reach a given level of capacity, based on what they have already done. The framework has been developed to guide organizational conversations on evaluation capacity and to provide more detail on what this entails – it is not meant to be used as a formal measure of capacity, although some organizations may wish to use it to develop a comprehensive, holistic picture of their evaluation capacity, as has been done in the third phase of this study.

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 study. This section will revisit key areas from the original literature review and highlight relevant findings for each of these. Once again, it should be noted that although the study focused specifically on evaluation capacity, most of the literature available in this field summarizes evaluation capacity building efforts. These two concepts (evaluation capacity and evaluation capacity building) have been separated here in an attempt to clarify the two in a manner reminiscent of Weber's ideal types (Weber, 1947). In reality, these are closely intertwined and factors influencing one can often be assumed to influence the other. This is why the following section revisits the original literature review, which focuses on ECB,

and establishes linkages between some of the key concepts found in ECB research and the study findings on evaluation capacity.

Evaluation Use

The utilization of evaluation (or evaluation use) was identified as a key consequence of evaluation capacity building in service organizations, especially in the context of participatory evaluation processes. The participation of stakeholders in the evaluation process is thought to lead to broader understandings of a program's theory and design, and to organizational learning about the program and its wider context (Patton, 1999; Preskill & Torres, 1994; Torres & Preskill, 2001; Turnbull, 1999). Even when evaluation methods are not fully participatory, as was the case in most of the organizations studied here, the literature points to a certain likelihood that an increase in evaluation capacity (to do and to use) would lead to increased evaluation use.

Several dimensions and sub-dimensions identified in the framework speak directly to evaluation use. The clearest example of this is the sixth dimension, Learning Benefits, which focuses entirely on use, with its instrumental/conceptual use (6.1) and process use (6.2) sub-dimensions. The descriptions used for each of the two sub-dimensions in this section of the framework summarize other published definitions of evaluation use, such as those by Alkin and Taut (2002), Johnson (1998), and Patton (1997). Some of the factors that have been found elsewhere to increase evaluation use have also made their way into the framework. For instance, Preskill, Zuckerman and Matthews (2003) have identified five categories of variables that influence process use. Some of these are found as sub-dimensions in the framework, such as "management support for evaluation", included under sub-dimension 3.5, Organizational Linkages, "frequency, methods, and quality of the communications occurring as part of the evaluation process" included under sub-dimensions 1.3, 3.3 and 3.5 (Communications and

Interpersonal Skills, Information Sharing, and Organizational Linkages, respectively), and “organizational stability and support for evaluation capacity building”, included under sub-dimension 2.3, Organizational Infrastructure. Other factors thought to influence evaluation use, such as the role of the evaluator (described in sub-dimension 3.2, Use of Consultants) and the involvement of program practitioners (described in sub-dimension 4.1, Involvement in Evaluation), were also identified in the literature on evaluation utilization and capacity (Alkin & Taut, 2002; Conner, 1998; Cousins, 1996; Toulemonde, 1999).

Interestingly enough, the framework contains very little information on participatory evaluation methods, even though this is thought to influence evaluation use to a significant extent. Although some participation from program stakeholders is included in the framework (e.g., in sub-dimension 4.1), most of the participants interviewed stated that in an effort to maintain neutrality and objectivity, their evaluation designs did not routinely include the participation of those closest to the object of evaluation. Some organizations put together steering committees or advisory boards for evaluation projects and collect feedback from program stakeholders through this mechanism. It is clear, however, that stakeholders can only comment on program information included in the various reports produced over the course of the study. Any comments on methodology or disagreements over recommendations are usually strongly discouraged by evaluation practitioners and left to the management response that follows an evaluation study. It seems, therefore, that it is possible to attain an exemplary level of capacity, at least on some dimensions, without espousing fully participatory evaluation practices.

The issue of the evaluator’s objectivity and neutrality echoes some of the inherent tensions between the learning and accountability functions of evaluation in the federal government. Although evaluation has always been used as an accountability mechanism, the

extent to which evaluation studies examined issues such as relevance and results has varied over the years. For a time, evaluation was thought to support program managers and was considered a learning tool more than an accountability mechanism (as described in Results for Canadians, 2001). In recent years, however, it is becoming increasingly clearer that evaluation should first and foremost meet the informational requirements of senior management, which focus more strongly on accountability issues. In fact, internal communications about the upcoming edition of the Evaluation Policy effective in April 2008 specify that the Deputy Minister is the key client of the evaluation function and that evaluations should focus on measuring the high-level socio-economic impacts of government programs, policies and initiatives.

This recent tendency to focus exclusively on accountability has significant implications for evaluation capacity building, especially in terms of its learning benefits and evaluation use. Under the current vision, program managers should be completely removed from the evaluation process and the evaluation report should be directed entirely towards the senior managers of the organization. In practice, however, this is not the case. Interview results show that evaluators still involve program managers and practitioners to a significant extent, and that they are concerned with making their evaluation studies relevant to program personnel as well as to senior management. This supports the view of evaluation use espoused in definitions of ECB and in the broader literature on evaluation utilization and stakeholder involvement.

Organizational Factors Contributing to the Success of ECB

Many of the factors identified in the evaluation literature as important to the success of ECB are found in the framework. It seems therefore that these may not only be precursors to successful capacity building activities as presented in the conceptual framework depicted in Chapter 2, but that they are a key part of evaluation capacity itself. As mentioned previously, the

two concepts are closely linked, and in fact, evaluation capacity can be thought of as a component of ECB (as per the conceptual framework developed in Chapter 2 and revisited later in this chapter). The four categories of organizational factors identified in the literature review (environment, structure, culture, and leadership) will be reviewed here and their relevance to the framework will be discussed in light of the study findings.

External Environment. The external environment of the organization influences the successful implementation of ECB initiatives through the demands for evaluation that it places on the organization, the incentives that it provides for conducting evaluation, and through the interactions that occur between external stakeholders and organizational members on matters of evaluation (Katz, Sutherland, & Earl, 2002; Mackay, 2002; Preskill et al., 2003; Stockdill et al., 2002; Sutherland, 2004; Toulemonde, 1999). The external environment of an organization has a clear influence on its evaluation capacity, as shown in the framework. For instance, sub-dimension 3.4 refers to the external supports available to evaluators, such as professional associations and other external networks. Evaluators are also frequently in contact with personnel from the Treasury Board on matters related to ongoing evaluation projects as well as to the broader evaluation policy that drives all evaluation activity within the government.

Organizational Structure. This category focuses primarily on the location of the evaluation function within the organization, its role, and its proximity to other key organizational areas, as outlined in sub-dimensions 3.2 and 3.5, Use of Consultants and Organizational Linkages (Preskill & Torres, 2000; Torres & Preskill, 2001). This section also discusses the presence of a centralized vs. decentralized evaluation function (Compton et al., 2002; Milstein et al., 2002, Sonnichsen, 1999), which echoes some of the comments made by one evaluation user who felt that decentralization allowed evaluators to gain a deeper understanding of the program

under study, without necessarily compromising their objectivity. The Organizational Structure² category also includes a discussion of the resources available for evaluative inquiry. These focus on the human and financial resources required to implement evaluation capacity building activities and to improve the state of evaluation within the organization in general (Bozzo, 2002; Cousins & Earl, 1992; King, 2002; Milstein et al., 2002; Stevenson et al., 2002; Stockdill et al., 2002; Trevisan, 2002). These are included in the framework, under dimension 1 (Human Resources) and sub-dimension 2.1 (Budget). In the framework, these focus on the conduct of evaluation studies and not on the implementation of ECB, as opposed to the literature, which focuses more heavily on the human and financial resources required to implement ECB initiatives.

Organizational Culture. This category of factors was found to be critical to the creation of a supportive organizational setting that influences individual and organizational learning and performance (Preskill, 1994; Preskill & Torres, 1999; Trevisan, 2002). Some of these factors are found in sub-dimensions 2.3 (Organizational Infrastructure) and 4.2 (Results-Management Orientation), which speak to the organizational climate and the importance given to evaluation at all levels of the organization. Sub-dimension 4.1 (Involvement in Evaluation) could also be included here, since the participation of stakeholders and their level of awareness and understanding of evaluation is key to fostering a learning climate within the organization (Boyle et al., 1999; Russ-Eft et al., 2002).

Organizational Leadership. A number of authors have focused on organizational leadership as a key factor in the successful implementation of ECB initiatives (Goh & Richards, 1997; Goh, 2003; King, 2002; Milstein et al., 2002). According to the literature reviewed, senior managers must make evaluative inquiry a priority and become advocates for learning in the

organization (Gibbs et al., 2002; Goh, 2003; Leonard & Miller, 1992; Preskill et al., 2003). Strong leadership was also found to be critical to evaluation capacity itself at several levels. It is addressed in the framework in sub-dimensions 1.5 (Leadership), 4.2 (Results-Management Orientation), and in dimension 5 (Integration with Organizational Decision-Making).

Stages of Evaluation Capacity Building

Four stages of ECB were derived from the literature and described in Chapter 2. These four stages are based on the steps taken by various organizations attempting to increase their evaluation capacity as related in a number of anecdotal reports on these efforts. The four stages do not describe the four capacity levels presented in the framework but represent the “behind the scenes” activities that result in organizations moving from one level to the next. This evolution does not necessarily always happen at the same time throughout the entire organization, or for all of the dimensions. Still, it may be useful to examine the four stages of evaluation capacity building (traditional evaluation, awareness and experimentation, implementing evaluative inquiry, and adoption of evaluation as a management function) in relation to the four levels of evaluation capacity (low, developing, intermediate, and exemplary), and to draw linkages between the two.

Traditional Evaluation. This first stage is characterized by the evaluation of individual programs, usually by external contractors specifically hired for this purpose (Gilliam et al., 2003; Sanders, 2002; Torres & Preskill, 2001). The main motivation behind evaluation at this stage is to meet external requirements, with little or no interest in organizational improvement or learning. This is usually the stage of ECB at which organizations with low evaluation capacity find themselves. Organizations with low evaluation capacity, according to the framework, have several vacant positions, project management issues, and little capacity in terms of technical

skills and interpersonal relationships. In addition to this, there is little impetus from management to collect performance data and to report on results in a rigorous fashion. Staff members are generally not aware of the need to conduct program evaluation, and there is little evidence that the evaluations produced are used to support decision-making or other organizational activities. These organizations also have no specific plans to develop their capacity to do or to use evaluation.

Awareness and Experimentation. The second stage begins when organizational members become aware of a need to change certain activities and experiment with different approaches to their way of functioning (Sanders, 2002; Torres & Preskill, 2001). These organizations could be said to have “developing” evaluation capacity, in that they recognize that there is value in building their evaluation capacity further, and start taking steps towards this goal. This level is characterized by difficulties in staffing the evaluation unit and ensuring that evaluators have the appropriate range of skills, limited leadership both in the evaluation unit and within the senior management ranks, scarce resources for evaluation, basic planning processes and little awareness of evaluation within the broader organizational context. Although some efforts are being made towards a working evaluation unit, it is still too early to speak of actual evaluation capacity within the organization.

Implementing Evaluative Inquiry. In this stage, the organization has experienced early successes from its evaluation activities and is looking to formalize it as a regular part of its functioning. Specific plans are developed and resources are committed to ECB initiatives (Duignan, 2003; Gibbs et al., 2002; Sanders, 2002; Torres & Preskill, 2001). This type of activity is manifested in organizations with intermediate evaluation capacity. These organizations have a full complement of highly skilled evaluators who work alongside more junior personnel

progressing along a pre-determined career path, have strong leadership within the evaluation unit and a supportive cast of senior managers who value and use evaluative information, collect ongoing performance data, allocate a stable budget to the evaluation function and show excellent evaluation planning. Other organizational members are aware and knowledgeable about evaluation's role and some are even involved directly in evaluation projects. There is some evidence of instrumental and conceptual use, and some process use is also thought to take place within the organization.

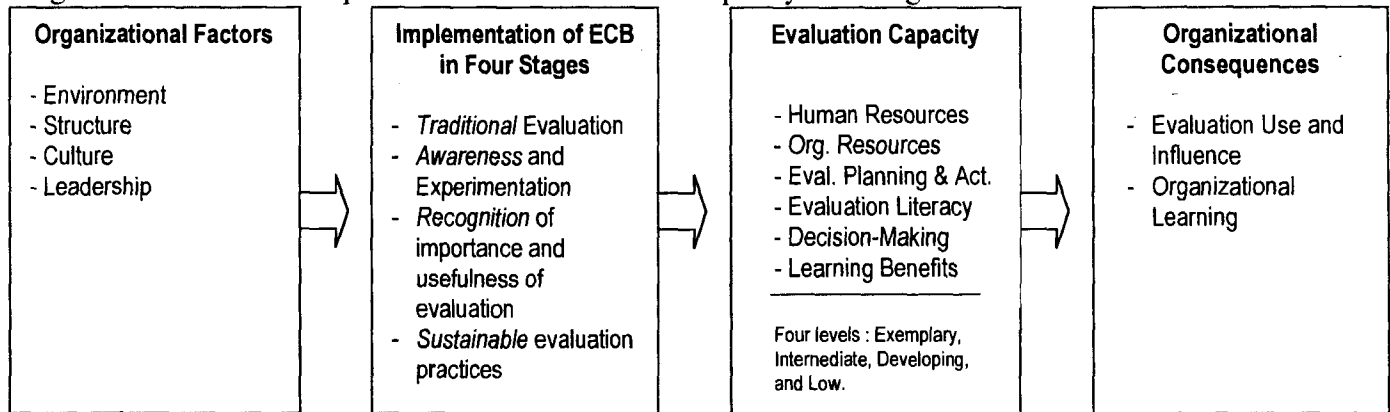
Adoption of Evaluation as a Management Function. This stage generally leads to the achievement of the "exemplary" capacity level found in the framework. Organizations that have reached this level of capacity value evaluative inquiry and consider it one of their core functions. New members are taught about evaluation and discussions about organizational values and achievements are ongoing (Gibbs et al., 2002; Lewis & Thornhill, 1994; Sanders, 2002; Torres & Preskill, 2001). Exemplary capacity in government organizations is characterized by a strong, highly-skilled evaluation function, quality evaluation products that are integrated seamlessly into organizational decision-making processes, excellent leaders within and outside the evaluation unit, a strong results-management orientation, a stable organizational structure and budget, a thorough planning process that involves a review of the evaluation unit itself, and clear evidence of evaluation utilization.

Implications for Conceptual Model of Evaluation Capacity

The conceptual model of evaluation capacity presented in Chapter 2 can be revisited in light of the findings of this study. Since the main goal of the study was to complete this conceptual model by identifying the main dimensions of evaluation capacity, the third box of the

model not previously filled in can now be completed. Figure 5 presents a revised conceptual model of evaluation capacity.

Figure 5: Revised Conceptual Model of Evaluation Capacity Building



Although much was known about the organizational factors necessary to the development of capacity and how this occurred over time, no empirically-derived data were previously available on the actual dimensions of evaluation capacity and how these would manifest in organizations. This study has provided the evidence necessary to complete the conceptual framework by developing a clear picture of evaluation capacity in government organizations. To revisit the logic model analogy used to describe the relationship between ECB and evaluation capacity used earlier, the revised conceptual model now clearly outlines the expected immediate and intermediate outcomes that should result from the implementation of ECB “activities and outputs”.

Other Important Findings

As presented in the previous sections, some elements of the evaluation capacity framework reach broader aspects of the conceptual framework shown above, such as the organizational factors required to ensure ECB success (e.g., leadership) and the consequences of ECB (e.g., evaluation use). At the same time, these elements can be taken at face value as

components of evaluation capacity itself, particularly within a governmental context. Other findings not directly related to the development of the framework are also informative and may point to a need for further research in this area. For example, organization size seems to be a significant factor in influencing evaluation capacity, although it is sometimes a positive force, sometimes a negative force. Interview data show that larger organizations often have the required resources and stable structure to implement evaluation capacity building, but experience issues related to reach and evaluation literacy. On the other hand, smaller organizations have better evaluation coverage and better linkages with senior management, but fewer resources to devote to complex evaluation projects and professional development.

Implications for Theory

The literature review presented in Chapter 2 outlines a theoretical critique of ECB based on Foucault's work. This critique will be revisited here and the findings of the study will be considered within this context. A discussion of some of the findings in relation to organizational micro-politics and institutional theory will also be presented. Finally, the implications of the findings for the organization growth theories explored at the outset of this study will be described in this section.

Governmentality and Evaluation. The governmentality approach derived from Foucault's work considers governmental institutions as organizations that direct the conduct of individual citizens by using a variety of incentives, demands and consequences delivered through their programs. The issue of power is critical to governmentality, since it is through its programs that government exerts an influence on the behaviour of individuals. Evaluation, within this context, ensures that the techniques (i.e., programs) used by government to shape the conduct of individuals lead to desired outcomes. As a "technology of performance" (Dean, 1999),

evaluation may miss the more important policy issues related to the relevance of government interventions in certain areas by focusing too closely on cost-effectiveness and program delivery. It appears as though the framework of evaluation capacity, as presented in Table 4, supports the status quo on two levels. First, the framework is based on the current model of a centralized evaluation function responding primarily to the informational needs of central agencies and senior managers. The framework does not account for alternative models used in some organizations, such as a decentralized evaluation function where individual evaluation specialists are employed by the programs themselves and coordinate evaluation activities at the program level. This alternative model is not used to a great extent in the Canadian federal government, mainly because of a concern for the perceived independence and objectivity of evaluators. This runs counter to the stance taken in the evaluation literature, which supports decentralization as a key ingredient in organizational learning (Caracelli, 2000; Jenlink, 1994; King, 2002; Milstein et al., 2002; Preskill, 1994; Sonnichsen, 1999). Second, by focusing on decision-making within a resource allocation context, the framework supports the current role espoused by evaluation in the federal government in terms of its propensity to focus on issues of cost-effectiveness and accountability rather than on larger, more complex issues related to the appropriateness of governmental intervention in certain social spheres.

In other words, although the framework supports the current centralized model of evaluation in the federal government, it also (and more importantly) brings to the forefront the key issues surrounding the utilization of evaluation by providing a systematic description of the different levels of evaluation use found in organizations. This is an important contribution to the current body of knowledge on evaluation capacity in particular, and has significant implications for ECB in organizations. By describing the levels of use of evaluation, the framework provides

a clear path for organizations interested in furthering their evaluation capacity in terms of learning and utilization. However, Foucault provides a conceptual frame within which evaluation use can be considered, especially in the current climate of accountability found throughout government. Foucault teaches us that concepts such as efficiency, found in most evaluations, are typically “treated as though they were neutral and technical matters, rather than being tied to particular interests. The question of ‘efficiency for whom’ is rarely asked. Efficiency itself is taken as self-evidently a good thing” (Ball, 1990, p. 154). The idea here is that the instrumental and conceptual uses of evaluation can sometimes lead to the blind support of the accepted managerial discourse without truly gaining a better understanding of the larger relevance issues that should be part of any evaluative activity. It may be helpful for organizations to consider the dimensions included in the framework under “capacity to use evaluation” in this light, to see through the direct application of the evaluation findings and recommendations in order to better understand the context within which these recommendations have been developed.

The fact that the use of the framework implies a certain measurement can also be viewed within the Foucauldian perspective. Despite caveats made at the outset of the study about the framework’s role as a description of evaluation capacity rather than an instrument, the application of the framework in four different organizations in the third phase of the study revealed that participants were able to use the framework to rate their organizations. Indeed, the inclusion of levels within the framework lends an implicit expectation of measurement that could have a significant impact on those who choose to use it in this manner. This has implications for performance appraisals and other possible disciplinary actions taken at the organizational level. Although the use of the framework by individuals is beyond the control of those who have

contributed to its development, further research work involving the framework should include a consideration of this issue in terms of study design and expected outcomes.

Disciplinary Power. Foucault's conception of disciplinary power, which applies to interpersonal relationships, has interesting implications on evaluation capacity building and on the framework developed in this study. Evaluators take on different roles in the course of their work; as described by Morabito (2000), these can include the role of educator, consultant, counselor, and facilitator. In each of these, the evaluator exerts power over other organizational members, especially in institutions where evaluation is primarily used for accountability purposes. These different roles and the power associated with them can be found throughout the framework. For instance, sub-dimension 4.1 (Involvement in Evaluation), has an element related to evaluation steering or advisory committees, which are used by evaluators to discuss evaluation projects with key stakeholders. The evaluator exerts some power over the individuals sitting on these committees, since he or she usually initiates and organizes committee meetings, decides of the agenda items to be discussed, and is typically recognized as the technical expert. Although program participants can discuss findings and comment on the evaluation methods used, their power is limited by the context of these meetings. Any comments they may have on the quality and methodological integrity of an evaluation are generally relegated to the management response produced at the end of an evaluation and any disagreements that may take place between evaluators and committee members are not recorded in the main evaluation report. The technical skills of evaluators (sub-dimension 1.2) and the extent to which the evaluators are perceived to be methodological experts also have a significant influence on the power relationship that develops between program managers and evaluators, as do their facilitation and communications skills during the evaluation process (sub-dimension 1.3). Once again, the

framework appears to support these power relationships by emphasizing the skills and abilities of evaluators (Dimension 1, Human Resources) over those of evaluation clients and stakeholders.

Finally, evidence of disciplinary power can be also drawn from the data collection process itself, without even considering the actual study findings. The different organizational roles adopted by the third phase participants likely had an impact on their responses during the interview. In all four organizations studied, the Heads of Evaluation and senior evaluators tended to be very positive about their organization's evaluation capacity and used positive language throughout the interview (i.e., "we're an exemplary organization", "we do most things very well", etc). In the three organizations where a decision-maker/evaluation user was interviewed, their comments were more balanced and they tended to provide more of a reality check, likely because they had less invested in the evaluation process than the other participants. Evaluation users tended to be more critical or more guarded in their comments for all six dimensions. This may point to the fact that evaluation personnel felt more "assessed" in their use of the framework than their evaluation user counterparts, since the framework dealt with issues directly related to the evaluation unit and its performance. It would be reasonable to think that the evaluators wanted to show their organization in the best possible light when using the framework, whereas evaluation users did not feel the need to do the same. This divergence of opinions and described experience is consistent with other studies conducted on evaluator and practitioner perspectives (Cousins, 2001). In this case, however, the power exerted by the framework itself is interesting to consider and might merit further consideration in future studies.

Micro-Politics. Traditional approaches to change in organizations treat change as an external process that is imposed upon people and organizations. The micro-political perspective instead views change as an integrated process that occurs within an organization and that is

heavily influenced by the actors in the organization (Ball, 1987). This perspective offers an interesting theoretical viewpoint through which some of this study's findings can be interpreted. This approach is similar to the Foucauldian perspective, in that it places the individuals involved within the organization and the power relationships that guide their interactions in the centre of the organizational change process:

Innovation arises in, and acquires legitimacy through group definitions and an appropriate structure of relationships. It is, therefore, a cultural phenomenon. Like any other idea-system, it is subject to the constraints of power distribution and the interactional opportunities which are available to the participating members. Even if the innovation originated at a time and place far removed from a specific institution, its realization inside that institution will nevertheless be mediated through its patterns of social interaction (Esland, 1972, p. 103).

The influence of micro-political processes in the comments made by participants in the third phase of the study is clear in many instances. For example, the rejection of the central evaluation function by Decision-Maker 2, interpreted under the light of micro-politics, may point to the fact that the use of this type of function poses a threat to the independence and self-sufficiency of the program areas. Decision-Makers 1 and 2 both spoke of the lack of understanding on the part of evaluators when it comes to details related to program design and function and how this raises potential concerns with the credibility and quality of recommendations made in evaluation reports. Producing recommendations, like other decision-making processes, can be seen as a micro-political process "which embraces a whole set of formal and informal arenas of interaction, confrontation and negotiation" (Ball, 1987, p. 237). It is not uncommon for stakeholders unhappy with a set of recommendations emanating from an

evaluation to discredit the methodology used in the evaluation or to cast the evaluators responsible for the report as biased or incompetent.

The extent to which organizations are able to develop their evaluation capacity can also be considered in terms of micro-political issues:

The introduction of, or proposal to introduce, changes in structure or working practices must be viewed in terms of its relationship to the immediate interests and concerns of those members likely to be affected, directly or indirectly. Innovations are rarely neutral. They tend to advance or enhance the position of certain groups and disadvantage or damage the position of others. Innovations can therefore threaten the self-interests of participants by undermining established identities, by deskilling and therefore reducing job satisfaction (Ball, 1987, p. 32).

The use of the framework in an organizational setting implies a certain measurement that can be fraught with micro-political issues. For example, the pride of the Heads of Evaluation was clearly visible in the Phase 3 interviews, especially in relation to the Human Resources and Evaluation Planning and Activities dimensions. These two dimensions are under the direct control of the Head of Evaluation, and this is where their actions have the most impact. The fact that all four Heads of Evaluation interviewed gave their organizations intermediate or exemplary ratings on these dimensions is not surprising in a micro-political perspective. Another example of micro-politics at work in the use of the framework is evident in Organization 4's request to add a level to the framework, to better reflect the difference between low and intermediate capacity. Although the two participants from this organization were uncomfortable for the most part with stating that their organization had achieved an intermediate level of capacity on the six dimensions, they also had issues with the use of the word "low" to describe their organization's

capacity. In their minds, low implied a complete lack of capacity, whereas “developing” could show early efforts in building evaluation capacity.

Organizations interested in further developing their evaluation capacity might do well to consider the micro-political implications of the innovations proposed. For example, a restructuring of the evaluation unit could cause some significant problems in terms of productivity. Along the same lines, the implementation of a performance measurement system or new requirements for RMAFs can result in certain problems related to the addition of new responsibilities to program staff, especially if the evaluation function retains a sign-off authority over the work that is done at the program level. In all of these cases, the potential ramifications of the proposed change should be taken into account by evaluators and senior managers.

Institutional Theory. Institutional theory is based on similar theoretical pillars as those of current organizational learning theory. In a nutshell, institutional theory rests on the basic idea that knowledge is socially constructed, in organizations as everywhere else, and is rooted in the seminal writings of Berger and Luckmann (1967) on constructivism. According to organizational learning theory, the diversity and breadth of the knowledge held by organizational members as well as how much this knowledge is shared with others are crucial to the ability of the organization to learn on a collective level (Coggshall, 2004). Organizational knowledge becomes embodied in routines, defined as the forms, rules, procedures, strategies, and technologies around which organizations are constructed. Organizations that have rules, norms and procedures that direct and constrain the behaviour of their members are thought of as institutions. Institutions seek legitimacy through three pillars (Scott, 2001): the regulative pillar, which comprises the formal and informal rules that regulate and control behaviour, the normative pillar, which is based on “normative rules that introduce a prescriptive, evaluative, and

obligatory dimension into social life” (Scott, 2001, p. 54), or the compliance to moral and ethical standards, and the cultural-cognitive pillar, which is composed of the “shared conceptions that constitute the nature of social reality and the frames through which meaning is made” (Scott, 2001, p. 57). Legitimacy is acquired through conforming to the rules that make up the regulative pillar, complying with the standards established in the normative pillar, and adopting a common frame of interpretation through the cultural-cognitive pillar, one that is similar to those that are found in society-at-large. According to institutional theory, organizations strive towards legitimacy in order to increase their share of resources from their environments and to enhance their competitiveness.

Much can be learned from institutional theory when applied to the question of evaluation capacity. Government departments and agencies derive their legitimacy from central agencies that have an oversight role as well as from the taxpayers that are an important part of their external environment. Accountability mechanisms such as evaluation can help organizations demonstrate that they meet regulatory standards and normative rules in a clear and direct manner. Evaluation questions pertaining to program design and cost-effectiveness typically focus on the program’s regulatory environment and on the intended and unintended impacts of the program on its stakeholders. Although it would be more challenging to use evaluation as a measure of the extent to which an organization meets the cultural-cognitive conceptions found in society at large, issues such as relevance are typically included in evaluation designs and can at least partially inform this issue. Evaluation therefore acts as an accountability mechanism, but can also be used as a communications tool to share information about the organization and achieve legitimacy in this manner.

The same theoretical framework can be applied to the study of evaluation capacity itself. The regulative, normative, and cultural-cognitive pillars can be conceptualized as forces that influence an organization's evaluation capacity. An example of regulative influences on an organization's evaluation capacity might be the Evaluation Policy and other policies that direct the work of evaluators. Evaluation capacity, under this pillar, would be described as the extent to which organizations meet the requirements set out by the Treasury Board Secretariat, such as the development and implementation of a five-year evaluation plan. Normative rules that may impact the evaluation capacity of an organization, for instance, could include informal requirements to seek out the comments of peer reviewers on evaluation reports, or the treatment of confidential information collected through key informant interviews. Although these are not formal rules established by Treasury Board, they are nonetheless embodiments of some of the standards of practice upheld by most evaluation associations. Evaluators in government, as elsewhere, generally feel that it is their professional responsibility to uphold these standards in their work. Finally, an example of evaluation capacity that is relevant to the cultural-cognitive pillar could be the development of RMAFs by program personnel. Organizations in which this occurs could be said to have a strong evaluation capacity, since most organizational members share a common frame of reference in terms of program outcomes and measures and are all working towards the same goal.

In addition to providing context for evaluation capacity building in an organization, institutional theory can also inform those interested in organizational problems or failures in building such capacity. For instance, organizations trying to increase the evaluation literacy of non-evaluators could study the interactions that are taking place between organizational members as well as the other forces at play, such as a heavy workload or a lack of general

awareness of the organization's management processes. This perspective can help organizational leaders take a step back and consider the issue within the broader frame of the organization and its environment.

Governmentality, micro-politics and institutional theory are all theoretical lenses through which evaluation capacity can be conceptualized. Each of these perspectives informs our understanding of the dimensions that compose evaluation capacity and allow us to think about each one in a deeper, more thorough manner. Although the primary purpose of this study was to identify and validate the dimensions of evaluation capacity, future research studies (such as those suggested later in this chapter) could base their methodological and epistemological approaches on one of these theories to gain a more complete understanding of how evaluation capacity actually contributes to organizational learning and evaluation use.

Profiling Organizational Growth. The process used to develop the evaluation capacity framework was borrowed from Leithwood and Montgomery's (1987) Innovation Profiles methodology. Conceptually, this approach features a process of "gap reduction", in which exemplary and low levels of an innovation are first identified, and the manageable steps required to move from one to the other are then developed to highlight the various milestones that can be expected to be reached along the way. Although these specific steps were taken by the researcher and not the participants themselves, the findings of the study point to the fact that the participants found the structure of the framework very intuitive and felt at ease with using the framework, especially once all of the dimensions and sub-dimensions had the same number of levels. The approach taken is therefore considered successful, and is recommended for future studies of a similar nature.

Limitations of the Study

The context within which this study was undertaken poses a number of limitations to the interpretation of its findings. These include the study's scope, the methodology selected for the study, and the interpretation of the framework by the study participants.

The study's unique focus on Canadian federal government organizations has generated findings that are applicable to these organizations in particular, but may not be appropriate for other types of organizations (e.g., non-profit, provincial government organizations, schools, etc.). Although a more specific focus has allowed for a deeper contextual analysis and increased the relevance of the framework for government organizations, some organizational features not found in the organizations studied may minimize the usefulness of the framework for non-governmental or other types of organizations.

The individual interview format selected for this study aimed to ensure that the participants would feel comfortable with using the framework behind closed doors, under the cover of confidentiality, as opposed to a focus group situation where they would be commenting on their organization's evaluation capacity in the presence of colleagues. This approach, however, may have yielded results that are very specific to the four organizations studied and may once again reduce the applicability of the framework across organizations, even within the federal government. This limitation was mitigated by selecting organizations with varying levels of evaluation capacity in order to obtain some level of representation of different organizational conditions.

An interesting aspect of the individual interview format, however, is the observed difference in responses to the Phase 3 interviews based on the organizational role of each participant. As could be expected, the Heads of Evaluation and senior evaluators were very

positive in their organizational assessments, while the evaluation users were somewhat more guarded in their enthusiasm. The evaluation professionals discussed the first three dimensions (i.e., “capacity to do”) at great length, while the evaluation users focused more specifically on the last three (i.e., “capacity to use”). This is likely based on each participant’s organizational role and personal experience with evaluation. The inclusion of evaluation users in the study therefore enhances the validity of the framework by ensuring that multiple points of view are represented in the dimensions of evaluation capacity.

The role of the researcher in the interviewing process may also have resulted in more positive assessments on the part of evaluation professionals, given that the researcher is an evaluation officer in a government agency. A certain level of competitiveness may have factored in the organizational application of the framework, or at least a concern with showing their organization in the best possible light. However, as the researcher knew many of the participants beforehand, this is not likely to have had a significant impact on the organizational assessments done by the third phase participants.

The small number of participating organizations has also resulted in some data loss, especially in the case of the low capacity organization, in which a suitable evaluation user could not be found to offer a balancing perspective to the assessment of the Head of Evaluation and senior evaluator. If more organizations had been invited to participate in the study, the loss of one participant out of 12 potential interviewees would have had a smaller impact on the body of data obtained. As it is, three evaluation users provided their views as part of the study instead of four. In addition to this, the particular view of the evaluation user in one of the two exemplary capacity organizations has resulted in further data loss because of this individual’s unwillingness to comment on the framework as it was presented. In the end, only two evaluation users provided

their views on the actual framework. However, it should be noted that the different perspective offered by this individual has added to the richness of the data collected in other ways by suggesting that the use of an alternative model to the traditional evaluation function could lead to different evaluation outcomes, as suggested in the literature by Sonnichsen (1999), who explains that a centralized evaluation function tends to support accountability requirements, while a decentralized function is preferable when the intended purpose of evaluation is formative assessment and improvement of programs.

The interpretation of the framework in the third phase of the study by participants also required the individuals interviewed to balance a willingness to provide an accurate picture of their organization with the temptation to show their organization in the best possible light. The differences between some of the comments made by evaluators and by their program counterparts may in part be explained by this difficulty. It can also be explained by the different points of view of the interviewees, based on their unique perspectives and roles. Further, the specificity of some of the elements found in the framework may make it difficult for some stakeholders to comment on the capacity level of their organization. For instance, some organizations have separate evaluation and performance management teams. In these organizations, it may be difficult for evaluators to comment on the status of RMAF development or on data quality because they are not directly involved in these processes. Along the same lines, it may be difficult for evaluation users to comment on the finer points of evaluation planning and budgeting, since this is far removed from their sphere of activity. Therefore, the framework should not be considered a ready-to-use tool for individual decision-makers, but a discussion aid that can be used within the context of broader organizational strategic planning and management activities. The framework could be used collectively, for instance, as a guide in

discussions between evaluation professionals and management during the evaluation planning process. It could also be used by Heads of Evaluation as they try to identify the skills and knowledge required to complete their team during staffing processes, or to highlight certain needed professional development activities in the development of learning plans or in the appraisal of the performance of employees.

Continuing Research

The work presented here addresses a key challenge associated with the study of evaluation capacity. By producing empirically-derived data on the dimensions and sub-dimensions of evaluation capacity, this study separates the concept of evaluation capacity from the broader issue of evaluation capacity building and outlines how evaluation capacity might appear in organizations. Now that the dimensions of evaluation capacity are known, new areas for research on this topic can be identified and potential research questions can be formulated.

Scope of Framework

As mentioned in the previous section, the framework developed as part of this study focuses specifically on Canadian federal government departments and agencies. Because of the differences that may exist between these and other types of organizations, future studies could investigate the applicability of the framework to other organizations in an effort to make the framework's scope broader. The validation of the framework in these other organizations could be conducted in the same way as was done in this study. A first step in this direction could be to validate the framework in other governmental organizations, such as provincial ministries or in other countries. Such an exercise could highlight the unique characteristics of the organizations used in this study and help identify other dimensions of evaluation capacity common to government organizations.

Another potential avenue for further research on evaluation capacity would be to identify alternative models for the evaluation function, such as the decentralized approach mentioned in one of the high capacity organizations participating in this study. Alternative approaches to evaluation might result in different definitions of evaluation capacity and might also have an impact on the scope of the framework. A first step in this area might be to identify those government organizations that have structured their evaluation function differently than the standard approach described in this study, and to apply the framework to assess its fit to these organizations' unique situations.

Finally, the ongoing sustainability of the framework should be revisited on a periodic basis to ensure the stability of the concepts of the framework over time and through changing contexts. The underlying character of evaluations in the federal government is often dependent upon the philosophy of the governing party, and is likely to change through the years, as new parties are elected to represent Canadians on Parliament Hill. Beyond the political landscape of evaluation in government, however, it is also likely that the framework will have to be adapted as organizations develop their capacity to do and use evaluation. Although the "exemplary" level is fairly broad and would be difficult to achieve across all dimensions, some aspects of the sub-dimensions may need modification over time to account for new organizational developments that move beyond what is currently described in the framework. It is also possible that the "low" capacity level will become obsolete over time and will need to be removed from the framework. Continuing research will be needed to ensure that the framework remains relevant in years to come.

Structure of Framework

The framework presented here is organized by dimension and sub-dimension of evaluation capacity. The decision to use such a structure stems from the Innovation Profiles methodology that was adapted to this study, as well as from the interview data obtained in the first phase. Continuing research on evaluation capacity could focus on a different structure in order to verify and further validate the framework's dimensions and sub-dimensions. For instance, one participant in the study suggested that there are three "pockets" of evaluation capacity in government organizations: the evaluation staff, the program managers, and the senior decision-makers. These three types of evaluation stakeholders could define the structure of a different framework of evaluation capacity, with dimensions that are directly tied to the roles of the individuals involved rather than to organizational characteristics. It would be interesting to see what differences would emerge and what dimensions would remain similar to those included in the current framework following such an exercise.

Alternative Research Methods

Qualitative individual interviews were used as a primary means of data collection in this study. Future validation exercises could use quantitative methods to test the framework in order to provide a different level of analysis. For instance, a survey questionnaire focusing on the dimensions and key dimensions of the framework, administered in a number of government organizations, could yield some interesting findings in terms of the dimensions considered most important and in terms of generating new dimensions or sub-dimensions. An interesting design to consider within this context is a matched sample design, where an evaluator and an evaluation user who have collaborated in the past would each complete a similar survey instrument. Their responses could be matched and analyzed in terms of their convergence and divergence for each

sub-dimension of the framework. Another potential method could involve a longitudinal study of evaluation capacity building in a government organization based on the framework. Such a study could use both quantitative and qualitative methods to draw a clear picture of the evolution of the organization's evaluation capacity. Finally, focus groups could be used to review and further validate the framework. Results from the focus groups could be compared to those obtained through the interviews in order to gauge whether major differences can be observed in the data generated by each method.

Other Development Models

One other avenue of research on evaluation capacity might be to review other developmental models used in government to identify the similarities and differences between these and the framework presented here. For example, the Management Accountability Framework (MAF), the Modern Comptrollership Capacity Framework, and others all speak of capacity building at some level. It would be interesting to see whether capacity in general can be summarized using the dimensions included in the framework (e.g., Organizational Resources could be applicable to various situations). Along the same lines, the Treasury Board's Centre of Excellence for Evaluation has recently released a Capacity Assessment Template (internal document) that includes questions related to the framework dimensions on Human Resources, Organizational Resources, Evaluation Planning and Activities, and to a limited extent, Integration with Organizational Decision-Making. It may be useful to see the similarities between the two documents, and where the evaluation capacity framework presented here might inform and improve the types of issues covered in the Capacity Assessment Template.

Organizational Change

One final area for continuing research worth considering is the place of the framework within the broader research work on organizational change. The structure of the evaluation capacity framework makes it an organizational change model where it is assumed that organizations will evolve towards a higher level of capacity based on their ECB efforts on certain dimensions. As can be seen in Rogers (1995), the adoption of organizational innovations occurs over time and through a number of steps. These steps can be characterized by resistance, dissonance, awareness-building, exploration, and discussion. Although the framework does not currently include these characteristics, the literature on the steps required to build evaluation capacity hints at such occurrences. Therefore, an interesting avenue for research might be to “zoom in” on the capacity building process within one specific sub-dimension and studying what happens as an organization moves from one level to the next.

Implications for Practice and Policy

In addition to opening up new possibilities for continuing research, the development of the evaluation capacity framework has certain implications for evaluation practice and policy, particularly in Canadian federal government organizations. The identification of the elements that compose evaluation capacity clarifies the types of activities that can be undertaken in order to increase an organization’s capacity to do and to use evaluation. Some recommendations for practice are provided in the paragraphs that follow, and have been organized according to these two categories. The last part of this section focuses on the use of the framework itself, and how its very existence may have implications for evaluation practice.

Capacity to Do Evaluation

A key issue in government organizations' capacity to do evaluation revolves around the availability of skilled personnel in the evaluation units. This issue poses a significant challenge throughout government for a number of reasons. First, the development of a skilled evaluator requires time and experience. The technical and interpersonal skills required to be a competent evaluator are developed over time, through the practice of evaluation and exposure to others' evaluation projects. As one participant stated, "you can't just crank them out of school". As a result, there are too few skilled evaluators to meet the increasing demand for evaluation services in departments, and Heads of Evaluation all over government find that they are "chasing the same talent". In addition to this, increasing retirement rates due to the aging of the public service and new resources from Treasury Board to hire evaluators (i.e., 90 new full-time evaluator positions will be created in 2008) are creating a shortage of senior-level evaluators capable of tackling complex evaluation projects and producing quality work. The elements identified in the "Staffing" sub-dimension of the Human Resources dimension of the framework point at some of the things that organizations can do to prevent or mitigate such a situation and ensure that quality evaluation work continues to be done, at least internally. For instance, the standardized career progression paths implemented by some exemplary capacity organizations have had a direct impact on staff retention and professional development.

Another key issue with respect to an organization's capacity to do evaluation is the leadership within the evaluation unit and at the senior levels of the organization. The importance given to evaluation and how it is supported and promoted, financially and otherwise, is key to the production of quality evaluation work and its dissemination across the organization. In other words, "building capacity requires activity at all levels" and this activity needs to be supported

by organizational leaders. In this respect, the framework provides some clues as to what needs to be done, such as providing continuing funding for evaluation (sub-dimension 2.1, budget) and implementing performance measurement systems that yield data suitable for evaluation purposes (sub-dimension 2.2, ongoing data collection).

The use of consultants (sub-dimension 3.2) was found to be a particularly contentious issue when discussing the framework with Heads of Evaluation and senior evaluators. In general, most participants agreed that a limited use of consultants in favour of a greater involvement of internal evaluators created higher evaluation capacity within the organization because of its effect on corporate memory and because internal evaluators often understand the complexities associated to the programs under study better than their private-sector counterparts. At the same time, an increasing concern for objectivity and neutrality on the part of evaluators within government and an effort to improve the transparency of program assessment efforts has resulted in calls for an increased role for external parties involved in evaluating government programs. The framework offers a combination of both perspectives by suggesting that organizations that possess exemplary evaluation capacity have developed an appropriate balance between the two, with evaluators fully engaged in projects (i.e., at a substantive level, not just as contract managers) and calling on consulting help when needed. This hybrid model offers interesting possibilities in terms of evaluation practice and policy within government, and strengthens the need for quality evaluation work conducted by consulting firms.

Capacity to Use Evaluation

The framework highlights the fact that many dimensions and sub-dimensions affect an organization's capacity to use evaluation. These dimensions and sub-dimensions may hint at potential areas for improvement in organizations wishing to increase their evaluation capacity.

How evaluation is used in an organization often sets the tone for how it is done and how it is disseminated. The “evaluation literacy” of organizational members (dimension 4), or what they perceive evaluation to be, plays an important role in evaluation use. As stated by one participant, “Evaluation is a term that continues to not be well understood; there is great focus on the word accountability because of the new act (i.e., Federal Accountability Act), and evaluation is not part of that discussion”. In most organizations, there is some degree of confusion between evaluation and audit, and evaluation and performance measurement. The overlap that exists between the spheres of influence of these three corporate areas can lead organizational members to mistake one for the other. Micro-politics, as described earlier in this chapter, might lend some interesting theoretical viewpoints through which to consider this issue. For example, demand for evaluation services, generally understood as a positive action within the context of evaluation capacity building, could be based on the fear of a potential audit rather than on an actual interest in the findings produced by evaluations.

Involvement in evaluation (sub-dimension 4.1) may help clarify the role and mandate of evaluators within the organization, and in turn, lead to greater demand for evaluation services (sub-dimension 5.1, management processes) and better use of evaluation results (sub-dimension 6.1, instrumental/conceptual use). In fact, the extent to which evaluators are involved in organizational activities such as the development of Memoranda to Cabinet or Treasury Board submissions is a clear indicator of an organization’s evaluation literacy and capacity to use evaluation.

The Learning Benefits dimension of the framework (dimension 6) offers clear guidance in terms of evaluation use that moves beyond senior-level decision making. The instrumental/conceptual use sub-dimension (6.1) outlines the various ways in which evaluation

findings and reports can be used to improve programs, while the process use sub-dimension (6.2) describes the broader impact of participation in and exposure to evaluation on organizational members and other stakeholders. Although it may be difficult for organizations to track these different types of uses, they have been included in the framework to promote evaluation use as a key outcome of the evaluation function in government.

Framework as Instrument of Evaluation Capacity

In the third phase of the study, the framework of evaluation capacity was validated by asking individuals in four different organizations to assess their organization based on the dimensions and sub-dimensions developed in the first and second phases. The purpose of this exercise was to identify missing elements and to verify the clarity of the wording used. Although the organizational assessment that was done in this phase was general and focused on validating the framework, several participants suggested that the framework would be a useful tool for them when making a case for increased evaluation resources or even just as a rating scale for their own evaluation capacity building activities. Participants expressed a clear interest in obtaining a final version of the framework for use in their organizations, and stated that it may be interesting to produce an organizational self-assessment tool based on a more quantitative measure of organizational evaluation capacity. A longer-term recommendation for both research and practice, therefore, is the transformation of the framework into an instrument that can be used by organizations to assess their level of evaluation capacity. This may involve the use of a rating scale for the various sub-dimensions or assigning a number of points per element or sub-dimension and per level. Such options could be considered in a future project and the instrument could be validated in much the same way as the framework was in this study. Broader methodological issues to be addressed at that time will include the reliability of such an

instrument in terms of its stability across raters, as well as the weightings of certain sub-dimensions based on their importance to the organization, as was done by Cousins et al. (2004).

The development of an instrument to measure organizational evaluation capacity does run certain inherent risks in terms of the power imbalances that its use may create. Issues such as who has responsibility for completing the assessment tool, who will be involved in completing it, and who will be held accountable for failure to progress annually, for example, will need some careful consideration. The instrument would be meant as a management aid in determining where organizations would most benefit from ECB initiatives and not as another control mechanism to be used either by central agencies or organizational senior managers. It may not be possible to reduce such risks, but they should at least be acknowledged and noted by those who wish to use the instrument.

Conclusion

The development of organizational capacity to evaluate programs, policies and initiatives and to use this information to improve organizational activity and to reach its far-reaching goals has been acknowledged repeatedly in the academic literature. The issue of evaluation capacity building, in particular, has generated a considerable amount of interest from evaluation scholars and practitioners, and has led to a substantial body of literature featuring anecdotal reports of ECB initiatives undertaken in different organizations. However, little empirical evidence has been produced that clearly identifies the features, or dimensions, of evaluation capacity in organizations. This study sought to fill this knowledge gap by providing answers to three research questions: What are the essential dimensions of evaluation capacity in Canadian federal government organizations? How are minimal and exemplary performance on each of these

dimensions characterized? What are the manageable steps required to move from minimal to exemplary performance?

The Canadian federal government was selected as the broader context for the study for theoretical and pragmatic reasons. First, the central agency governing the evaluation function within government, the Treasury Board Secretariat, has given significant thought to evaluation capacity building in recent years as it deals with increased accountability pressures from taxpayers and senior government executives. This has created increased awareness on the part of departments and agencies to enhance their capacity to produce quality evaluation and to increase their ability to use evaluation findings in decision-making processes. Although some organizations within the federal government have dealt with this issue for a number of years and have developed their evaluation capacity significantly, others are just starting to make the change from activity-focused management to results-based management. The differences between these organizations set an interesting context for a study of evaluation capacity. At a more practical level, the availability and accessibility of departments and agencies willing and able to participate in the study underscored the usefulness of studying evaluation capacity in the federal government.

A three-phased approach was used to generate answers to the three evaluation questions. The first phase focused on identifying the key dimensions of evaluation capacity through an in-depth literature review followed by a series of semi-structured interviews with evaluation experts. Four individuals were selected for this first phase, based on their exposure to government evaluation activities and their broad view of evaluation in government. Two of these participants were experienced current or former consultants working on evaluation-related projects for government clients. The other two participants were current or former high-level

officials working on government evaluation policy on behalf of the Treasury Board Secretariat. The literature review and interview findings were summarized in a draft framework of evaluation capacity. The four experts were also asked to suggest federal government organizations at a low, intermediate, and high level of evaluation capacity, for later use in the study.

The second phase of the study sought to refine the framework by conducting a second interview with the four experts consulted in the first phase. This time, experts were asked to review the framework and provide feedback on its contents and wording. Their comments were integrated into a revised framework.

The third phase of the study focused on validating the revised framework in four government departments and agencies. The organizations were selected based on the recommendations of the four experts consulted in the first two phases of the study. Five organizations were originally selected; however, one was unable to participate due to turnover issues. Three individuals in each of the remaining four organizations were contacted and asked if they were willing to participate in the study: the head of evaluation, a senior evaluator, and an evaluation user. Eleven individuals agreed to participate. They were asked to review the draft framework and to provide feedback on its contents and wording. The feedback obtained through the interviews was used to finalize the framework of evaluation capacity.

This study concluded that evaluation capacity in Canadian federal government departments and agencies can be described functionally and operationally through six main dimensions that reflect an organization's ability to do evaluation and its ability to use evaluation: human resources, organizational resources, evaluation planning and activities, evaluation literacy, organizational decision-making, and learning benefits. Each of these dimensions was further broken down into a number of sub-dimensions. The capacity of these organizations on

each of these dimensions and sub-dimensions can be assessed using four levels: low, developing, intermediate, and exemplary. The study found that organizations vary in terms of their capacity from one dimension to the next, and indeed, from one sub-dimension to the next. In other words, an organization can have an exemplary level in terms of its staffing practices, but can be assessed as having low capacity in terms of its results-management orientation. This demonstrates that the framework is useful in helping organizations pinpoint areas in which they may have weaker capacity as well as the manageable steps required to increase their capacity.

This study makes an important contribution to the body of knowledge on evaluation capacity; although much has been published in the evaluation literature on evaluation capacity building, the actual characteristics and attributes of evaluation capacity itself have not before been defined and described based on empirical data. In addition, the study yields important clues as to what a theory of change on evaluation capacity might look like. The framework not only provides a systematic description of a complex organizational phenomenon; it also advances our thinking on organizational development and improvement. Continuing research may focus on expanding the scope of the framework to other types of organizations, while the main practical implication of this study is the potential use of the framework as an instrument of evaluation capacity for government organizations. Given that the Treasury Board Secretariat is poised to release a new Evaluation Policy that is rumoured to require complete evaluation coverage of a department's programs within a five-year timeframe, questions of evaluation capacity are likely to take centre stage in discussions held between departments and central agencies. Added to current concerns on the part of a national evaluation organization about the credentialing of evaluation professionals, such issues may create a demand for more empirical data about evaluation capacity in government and other types of organizations. The current study provides

useful background information to inform the discussions taking place as well as concise empirical data that can be used as a launch pad for new research work.

REFERENCES

- Adorno, T.W. (2000). **Introduction to Sociology**, trans. E. Jephcott. Stanford, CA: Stanford University Press.
- Alkin, M. C., and Taut, S.M. (2002). Unbundling evaluation use. **Studies in Educational Evaluation**, **29**, 1-12.
- Argyris, C., and Schon, D. (1978). **Organizational Learning: A Theory of Action Perspective**. Reading, Mass: Addison-Wesley.
- Baizerman, M., Compton, D.W., and Stockdill, S.H. (2002). Summary and analysis of the case studies: Themes across the cases. In M. Baizerman, D.W. Compton, & S.H. Stockdill (Eds.), **The art, craft and science of evaluation capacity building. New Directions for Evaluation**, No. 93 (pp. 101-107). San Francisco: Jossey-Bass.
- Ball, S.J. (1987). **The Micro-Politics of the School: Towards a Theory of School Organization**. London: Methuen.
- Berger, P.L. and Luckmann, T. (1967). **The Social Construction of Reality: A Treatise in the Sociology of Knowledge**. Toronto: Random House.
- Berends, H., Boersma, K., and Weggeman, M. (2003). The structuration of organizational learning. **Human Relations**, **56**(9), 1035-1056.
- Boyle, R., Lemaire, D., and Rist, R.C. (1999). Introduction: Building evaluation capacity. In R. Boyle and D. Lemaire (Eds.), **Building Effective Evaluation Capacity: Lessons from Practice** (pp. 1-19). New Brunswick, NJ: Transaction Publishers.
- Bozzo, S. L. (2002). Evaluation capacity building in the voluntary/nonprofit sector. **Canadian Journal of Program Evaluation**, **17**, 75-92.
- Brookfield, S. (2001). Unmasking power: Foucault and adult learning. **Canadian Journal for Studies in Adult Education**, **15**, 1-23.
- Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. In V. J. Caracelli and H. Preskill (Eds.), **The expanding scope of evaluation use. New Directions for Evaluation**, No. 88 (pp. 99-111). San Francisco: Jossey-Bass.
- Carnall, C.A. (2003). **Managing Change in Organizations** (4th Ed.). Essex: Pearson Education Ltd.
- Christie, C.A. (2007). Reported influence of evaluation data on decision makers' actions: An empirical examination. **American Journal of Evaluation**, **28**(1), 8-25.

- Colebatch, H.K. (2002). Government and governmentality: Using multiple approaches to the analysis of government. **Australian Journal of Political Science**, *37* (3), 417-435.
- Coggshall, J.G. (2004). Reform refractions: Organizational perspectives on standards-based reform. In Wayne K. Hoy and Cecil G. Miskel (Eds.), **Educational Administration, Policy and Reform: Research and Measurement** (pp.1-23). Greenwich, CT: Information Age Publishing.
- Comfort, L. K. (1997). Toward a theory of transition in complex systems. **American Behavioral Scientist**, *40*(3), 375-383.
- Compton, D. W., Glover-Kudon, R., Smith, I.E., and Avery, M.E. (2002). Ongoing capacity-building in the American Cancer Society (ACS) 1995-2001. In M. Baizerman, D.W. Compton, and S.H. Stockdill (Eds.), **The art, craft, and science of evaluation capacity building. New Directions for Evaluation, No. 99** (pp. 47-61). San Francisco: Jossey-Bass.
- Conner, R.F. (1998). Toward a social ecological view of evaluation use. **American Journal of Evaluation**, *19* (2), 237-241.
- Cooksy, L.J. & Caracelli, V.J. (2005). Quality, context and use: Issues in achieving the goals of metaevaluation. **American Journal of Evaluation**, *26*(1), 31-42.
- Cousins, J. B. (1996). Consequences of researcher involvement in participatory evaluation. **Studies in Educational Evaluation**, *22*, 3-27.
- Cousins, J.B. (2001). Do evaluator and practitioner perspectives converge in collaborative evaluation? **Canadian Journal of Program Evaluation**, *16*(2), 113-133.
- Cousins, J.B., Aubry, T.D., Smith Fowler, H., and Smith, M. (2004). Using key component profiles for the evaluation of program implementation in intensive mental health case management. **Evaluation and Program Planning**, *27* (1), 1-23.
- Cousins, J. B., and Earl, L.M. (1992). The case for participatory evaluation. **Educational Evaluation and Policy Analysis**, *14*, 397-418.
- Cousins, J. B., Goh, S.C., Clark, S., and 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*, 99-141.
- Cousins, J.B., and Leithwood, K.A. (1993). Enhancing knowledge utilization as a strategy for school improvement. **Knowledge: Creation, Diffusion, Utilization**, *14*, 305-333.
- Cousins, J.B., and Whitmore, E. (1998). Conceptualizing participatory evaluation. In E. Whitmore (Ed.), **Participatory Evaluation Approaches. New Directions for Evaluation, No. 88** (pp. 5-23). San Francisco: Jossey-Bass.

- Creswell, J. (1998). **Qualitative Inquiry and Research Design; Choosing Among Five Traditions**. Thousand Oaks, CA: Sage Publications.
- Dean, Mitchell (1999). **Governmentality: Power and Rule in Modern Society**. Thousand Oaks, CA: Sage Publications.
- DiBella, A. J., and Nevis, E.C. (1998). **How Organizations Learn: An Integrated Strategy for Building Learning Capacity**. San Francisco: Jossey-Bass.
- DiMaggio, P.J., and Powell, W.W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. **American Sociological Review**, 10 (4), 803-813.
- Duignan, P. (2003). Mainstreaming evaluation or building evaluation capability? Three key elements. In J. R. Sanders (Ed.), **The mainstreaming of evaluation. New Directions for Evaluation, No. 99** (pp. 7-21). San Francisco: Jossey-Bass.
- Fletcher, B. R. (1990). **Organization Transformation Theorists and Practitioners: People and Themes**. New York, NY: Praeger.
- Forss, K., Cracknell, B., and Samset, K. (1994). Can evaluation help an organization to learn? **Evaluation Review**, 18 (5), 574-591.
- Franckreiss, A., and Victor, P. (2002). The five dimensions of change: An integrated approach to strategic organizational change management. **Strategic Change**, 11, 35-42.
- Fullan, M. (1999). **Change Forces: The Sequel**. London: Falmer Press.
- Gibbs, D., Napp, D., Jolly, D., Westover, B., and Uhl, G. (2002). Increasing evaluation capacity within community-based HIV prevention programs. **Evaluation and Program Planning**, 25, 261-269.
- Gilliam, A., Barrington, R., Davis, D., Lacson, R., Uhl, G., and Phoenix, U. (2003). Building evaluation capacity for HIV prevention programs. **Evaluation and Program Planning**, 26, 133-142.
- Goh, S. (2003). Improving organizational learning capability: Lessons from two case studies. **Learning Organization**, 10, 216-227.
- Goh, S., and Richards, G. (1997). Benchmarking the learning capability of organizations. **European Management Journal**, 15(5), 575-583.
- Government of Canada, Treasury Board Secretariat (2000). **Results for Canadians**. Document retrieved at http://www.tbs-sct.gc.ca/report/res_can/rc_1_e.asp#pre on January 5th, 2006.

- Government of Canada, Treasury Board Secretariat (2001). **Evaluation Policy**; retrieved February 3rd, 2008 from: http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/tbm_161/ep-pe_e.asp.
- Government of Canada, Treasury Board Secretariat (2002). **Competency Profiles for Federal Public Service Evaluation Professionals**. Document retrieved on July 25th, 2007, from: at http://www.tbs-sct.gc.ca/eval/stud_etud/capa-pote02_e.asp.
- Government of Canada, Treasury Board Secretariat (2004). **Evaluation Function in the Government of Canada (Draft)**, retrieved February 3rd, 2008 from: http://www.tbs-sct.gc.ca/eval/pubs/func-fonc/func-fonc_e.asp.
- Government of Canada, Treasury Board Secretariat (2007). **Evaluation Policy Renewal**, retrieved on February 3rd, 2008 from: http://www.tbs-sct.gc.ca/eval/tools_ouils/renou_pol_renew/renou_pol_renew_e.asp
- Greene, J.C. (2002). With a splash of soda, please: Towards active engagement with a difference. **Evaluation**, **8**, 259-266.
- Hall, G. E., Loucks, S.F., Rutherford, W.L., and Newlove, B.W. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. **Journal of Teacher Education**, **26**, 5-10.
- Hannay, L.M., and Ross, J. A. (1997) Initiating secondary school reform: The dynamic relationship between restructuring, reculturing, and retiming. **Educational Administration Quarterly**, **33**, 576-603.
- Henry, G.T., and Mark, M.M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. **American Journal of Evaluation**, **24**(3), 293-314.
- Hill, F. M., Hazlett, S., and Meegan, S. (2001). A study of the transition from ISO 9000 to TQM in the context of organizational learning. **International Journal of Quality and Reliability Management**, **18**(2), 142-168.
- Horton, D. et al. (2003). **Evaluating Capacity Development: Experiences from Research and Development Organizations Around the World**. Ottawa, ON: IDRC.
- Howe, K.R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. **Educational Researcher**, **17** (8), 10-16.
- Huber, G.P. (1991). Organizational learning: The contributing processes and the literatures. **Organization Science**, **2** (1), 88-115.
- Huberman, A.M. (1994). Research utilization: The state of the art. **Knowledge & Policy**, **7** (4), 13-33.

- Inheteen, K. (1999). Can gender equality be institutionalized? The role of launching values in institutional innovation. **International Sociology**, 14(4), 403-422.
- Jenlink, P. M. (1994). Using evaluation to understand the learning architecture of an organization. **Evaluation and Program Planning**, 17, 315-325.
- Johnson, J. R. (1998). Embracing change: A leadership model for the learning organization. **International Journal of Training and Development**, 2(2), 141-150.
- Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. **Evaluation and Program Planning**, 21, 93-110.
- Jolly, D., Gibbs, D., Napp, D., Westover, B., & Uhl, B. (2003). Technical assistance for the evaluation of community-based HIV prevention programs. **Health Education and Behavior**, 30 (5), 550-563.
- Jorjani, H. (1998). Demystifying results-based performance measurement. **Canadian Journal of Program Evaluation**, 13, 61-95.
- Katz, S., Sutherland, S., and Earl, L. (2002). Developing an evaluation habit of mind. **Canadian Journal of Program Evaluation**, 17, 103-119.
- Kezar, A. J. (2001). **Understanding and Facilitating Organizational Change in the 21st Century: Recent Research and Conceptualizations** (Volume 28, Number 4). New York, NY.
- King, J. A. (2002). Building the evaluation capacity of a school district. In M. Baizerman, D.W. Compton, and S.H. Stockdill (Eds.), **The art, craft, and science of evaluation capacity building. New Directions for Evaluation, No. 99** (pp. 63-80). San Francisco: Jossey-Bass.
- Kirkhart, K.E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In V.J. Caracelli and H. Preskill (Eds.), **The expanding scope of evaluation use. New Directions for Evaluation, No. 88** (pp. 5-23). San Francisco: Jossey-Bass.
- Klimecki, R., and Lasseben, H. (1998). Modes of organizational learning: Indications from an empirical study. **Management Learning**, 29(4), 405-430.
- Kruse, S. D. (2003). Remembering as organizational memory. **Journal of Educational Administration**, 41, 332-347.
- Leithwood, K.A., and Montgomery, D.J. (1987). **Improving Classroom Practice Using Innovation Profiles**. Toronto: OISE.
- Leonard, E. C., and Miller, M.J. (1992). The role of evaluation in increasing the usefulness of employee involvement programs. **Evaluation Practice**, 13(1), 9-13.

- Leviton, L.C. (2003). Evaluation use: Advances, challenges and applications. **American Journal of Evaluation**, 24 (4), 525-535.
- Lewis, P. and Thornhill, A (1994). The evaluation of training: An organizational culture approach. **Journal of European Industrial Training**, 18 (8), 25-32.
- Lincoln, Y.S., and Guba, E.G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. **New Directions for Program Evaluation**, No. 30, 73-84.
- Loucks-Horsley, S., and Hergert, L.F. (1985). **An Action Guide to School Improvement**. Andover, MA: ACSD.
- Lunenburg, F.C., and Ornstein, A.C. (2004). **Educational Administration: Concepts and Practices** (4th Ed.). Toronto: Thomson.
- Lusthaus, C., Adrien, M., Anderson, G., Carden, F., and Montalvan, G.P. (2002). **Organizational Assessment: A Framework for Improving Performance**. Ottawa, ON: IDRC.
- Mackay, K. (2002). The World Bank's ECB experience. In M. Baizerman, D.W. Compton, and S.H. Stockdill (Eds.), **The art, craft, and science of evaluation capacity building**. **New Directions for Evaluation**, No. 99 (pp. 81-99). San Francisco: Jossey-Bass.
- Mark, M.M., and Henry, G.T. (2004). The mechanisms and outcomes of evaluation influence. **Evaluation**, 10, 35-57.
- Matlay, H. (2000). Organizational learning in small learning organizations: An empirical overview. **Education and Training**, 42(4/5), 202-210.
- Miller, T.I., Kobayashi, M.M., and Noble, D.M. (2006). Insourcing, not capacity building, a better model for sustained program evaluation. **American Journal of Evaluation**, 27 (1), 83-94.
- Miller, D., Lant, T.K., Milliken, F.J., and Korn, H.J. (1996). The evolution of strategic simplicity: Exploring two models of organizational adaptation. **Journal of Management**, 22 (6), 863-887.
- Milstein, B., Chapel, T.J., Wetterhall, S.F., and Cotton, D.A. (2002). Building capacity for program evaluation at the centers for disease control and prevention. In M. Baizerman, D.W. Compton, and S.H. Stockdill (Eds.), **The art, craft, and science of evaluation capacity building**. **New Directions for Evaluation**, No. 99 (pp. 27-46). San Francisco: Jossey-Bass.
- Morabito, S. M. (2002). Evaluator roles and strategies for expanding evaluation process influence. **American Journal of Evaluation**, 23, 321-330.

- Owen, J. M., and Lambert, F.C. (1998). Evaluation and the information needs of organizational leaders. **American Journal of Evaluation**, 19(3), 355-365.
- Newcomer, K. E. (2004). How might we strengthen evaluation capacity to manage evaluation contracts? **American Journal of Evaluation**, 25(2), 209-218.
- Patton, M.Q. (1997). **Utilization-focused Evaluation: The New Century Text**. Thousand Oaks, CA: Sage Publications.
- Patton, M. Q. (1999). Organizational development and evaluation. **Canadian Journal of Program Evaluation**, Special Issue, 93-113.
- Payne, S. L. (1996). Qualitative research and reflexive faculty change potentials. **Journal of Organizational Change Management**, 9(2), 20-32.
- Phillips, D.C. (1990). Postpositivistic science: Myths and realities. In E.G. Guba (Ed.), **The Paradigm Dialog**. Newbury Park, Ca: Sage Publications.
- Poole, W.L. (1996). The teacher unions' role in 1990s educational reform: An organizational evolution perspective. **Educational Administration Quarterly**, 37 (2), 173-196.
- Popper, M. and Lipshitz, R. (2000). Organizational learning: Mechanisms, culture, and feasibility. **Management Learning**, 31 (2), 181-196.
- Porteous, N.L. & Sheldrick, B.J. (1999). Enhancing managers' evaluation capacity: A case study from Ontario public health. **Canadian Journal of Program Evaluation, Special Issue**, 137-154.
- Preskill, H. (1994). Evaluation's role in enhancing organizational learning: A model for practice. **Evaluation and Program Planning**, 17, 291-297.
- Preskill, H., and Torres, R.T. (1994). Evaluation's role in enhancing organizational learning: A model for practice. **Evaluation and Program Planning**, 17, 291-297.
- Preskill, H., and Torres, R.T. (1999). **Evaluative Inquiry for Learning in Organizations**. Thousand Oaks, CA: Sage Publications.
- Preskill, H., and Torres, R.T. (2000). The learning dimension of evaluation use. In V. J. Caracelli and H. Preskill (Eds.), **The expanding scope of evaluation use. New Directions for Evaluation, No. 88** (pp. 25-36). San Francisco: Jossey-Bass.
- Preskill, H., Zuckerman, B., and Matthews, B. (2003). An exploratory study of process use: Findings and implications for future research. **American Journal of Evaluation**, 24, 423-442.

- Prochaska, J. M., Prochaska, J.O., and Levesque, D.A. (2001). A transtheoretical approach to changing organizations. **Administration and Policy in Mental Health**, **28**(4), 247-262.
- Rajagopalan, N., and Spreitzer, G.M. (1997). Toward a theory of strategic change: A multi-lens perspective and integrative framework. **Academy of Management Review**, **22**(1), 48-79.
- Robinson, T. T., and Cousins, J.B. (2004). Internal participatory evaluation as an organizational learning system: A longitudinal case study. **Studies in Educational Evaluation**, **30**, 1-22.
- Rogers, E. M. (1995). **Diffusion of Innovations** (4th ed.). New York: Free Press.
- Rogers, P. J., and Hough, G. (1995). Improving the effectiveness of evaluations: Making the link to organizational theory. **Evaluation and Program Planning**, **18**, 321-332.
- Rose, N., & Miller, P. (1992). Political power beyond the state: Problematics of government. **British Journal of Sociology**, **43** (2), 173-205.
- Rowe, W. E., and Jacobs, N.F. (1998). Principles and practices of organizationally integrated evaluation. **Canadian Journal of Program Evaluation**, **13**(1), 115-138.
- Russ-Eft, D., Atwood, R., and Egherman, T. (2002). Use and non-use of evaluation results: Case study of environmental influences in the private sector. **American Journal of Evaluation**, **23**, 19-31.
- Sanders, J. R. (2002). Presidential address: On mainstreaming evaluation. **American Journal of Evaluation**, **23**, 253-259.
- Sanders, J.R. (2003). Mainstreaming Evaluation. In J.J. Barnette & J.R. Sanders (Eds.), **The mainstreaming of evaluation. New Directions for Evaluation**, No. **99** (pp. 3-6). San Francisco: Jossey-Bass.
- Scott, W.R. (2001). **Institutions and Organizations** (2nd Ed.). Thousand Oaks, CA: Sage.
- Scott, W.R. (2003). **Organizations: Rational, Natural, and Open Systems** (5th Ed.). Upper Saddle River, NJ: Prentice Hall.
- Segsworth, R. (2005). Program evaluation in the government of Canada: Plus ça change...**Canadian Journal of Program Evaluation**, **20** (3), 175-197.
- Shulha, L.M., and Cousins, J.B. (1997). Evaluation use: Theory, research, and practice since 1986. **Evaluation Practice**, **18** (3), 195-208.

- Sonnichsen, R. C. (1999). Building evaluation capacity within organizations. In R. Boyle and D. Lemaire (Eds.), **Building Effective Evaluation Capacity: Lessons From Practice** (pp. 53-73). New Brunswick, NJ: Transaction Publishers.
- Stevens, J.A., and Slack, T. (1998). Integrating social action and structural constraints: Towards a more holistic explanation of organizational change. **International Review for the Sociology of Sport**, **33** (2), 143-154.
- Stevenson, J. F., Florin, P., Mills, D.S., and Andrade, M. (2002). Building evaluation capacity in human service organizations: A case study. **Evaluation and Program Planning**, **25**(2), 233-243.
- Stockdill, S. H., Baizerman, M., and Compton, D. (2002). Toward a definition of the ecb process: A conversation with the ecb literature. In M. Baizerman, D.W. Compton, & S.H.
- Stockdill (Eds.), **The art, craft and science of evaluation capacity building. New Directions for Evaluation, No. 93** (pp. 7-25). San Francisco: Jossey-Bass.
- Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison project school. **American Journal of Evaluation**, **25**, 277-293.
- Taylor, F.W. (1947). **The Principles of Scientific Management**. New York: Harper & Brothers.
- Torres, R., and Preskill, H. (1999). Ethical dimensions of stakeholder participation and evaluation use. **New Directions for Evaluation**, **82**, 57-66.
- Torres, R., and T. Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. **American Journal of Evaluation**, **22**, 387-395.
- Toulemonde, J. (1999). Incentives, constraints and culture-building as instruments for the development of evaluation demand. In R. Boyle. and D. Lemaire (Eds.), **Building Effective Evaluation Capacity: Lessons from Practice** (pp. 153-177). New Brunswick, NJ: Transaction Publishers.
- Trevisan, M. S. (2002). Evaluation capacity in K-12 school counseling programs. **American Journal of Evaluation**, 291-305.
- Turnbull, B. (1999). The mediating effect of participation efficacy on evaluation use. **Evaluation and Program Planning**, **22**, 131-140.
- Vakola, M. (2000). Exploring the relationship between the use of evaluation in business process re-engineering and organizational learning and innovation. **Journal of Management Development**, **19**(10), 812-835.

- Van de Ven, A. H. (1993). Managing the process of organizational innovation. In G. P. Huber and W. H. Glick (Eds.), **Organizational Change and Redesign** (pp. 269-294). New York: Oxford University Press.
- Van de Ven, A. H., and Poole, M.S. (1995). Explaining development and change in organizations. **Academy of Management Review**, **20**(3), 510-540.
- Weber, M. (1947). **The Theory of Social and Economic Organizations**, trans. Talcott Parsons. New York: Oxford University Press.
- Weiss, C.H., Murphy-Graham, E., and Birkeland, S. (2005). An alternate route to policy influence: How evaluations affect D.A.R.E. **American Journal of Evaluation**, **26**(1), 12-30.
- Wenger, E. (2004). Communities of practice and social learning systems. In S. T. a. A. M. K. Starkey (Ed.), **How Organizations Learn** (pp. 238-258). London: Thomson Learning.
- Yeung, A.K., Ulrich, D.O., Nason, S.W., and Von Glinow, M.A. (1999). **Organizational Learning Capability**. New York, NY: Oxford University Press.

APPENDIX A

Phase 1 - Introductory Email

Dear _____,

I am currently conducting a research project on evaluation capacity in the Canadian federal government as part of my Ph.D. thesis in measurement and evaluation at the Faculty of Education, University of Ottawa. Part of my research work involves conducting interviews with people working in the field of evaluation in Ottawa and who have been involved with a number of federal government departments and agencies through their work and other professional activities. As (*describe person's role here*), I am seeking your participation in two of these interviews, conducted at different times in the research process.

Each interview should take approximately one hour and will be held in person, at your place of business or other location of your choice. Please be assured that the information 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 which 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.

If you are willing to participate in these interviews, please reply to this email indicating when you are available over the next few weeks or call _____) to speak with me directly. I would be happy to answer any questions on the research project or the interview prior to our first meeting, or when we meet. You may also address questions or concerns to Professor J. Bradley Cousins at the University of Ottawa, who is supervising this thesis (613.562.5800 ext. 4036 or bcousins@uottawa.ca).

Thank you very much for taking the time to consider this request. I look forward to hearing your insights on this interesting topic.

Sincerely,

Isabelle Bourgeois
Ph.D. Candidate, Faculty of Education
University of Ottawa

APPENDIX B

Phase 1 - Letter of Informed Consent

I agree to participate in the doctoral research project conducted by Isabelle Bourgeois at the Faculty of Education, University of Ottawa. My participation involves two in-person interviews, each taking approximately one hour, with the researcher on the evaluation function within the Government of Canada. The second interview will also require the review of a framework that identifies the dimensions of evaluation capacity in federal government organizations. Both interviews will be conducted at my place of work or another public location of my choice. I understand that the information I will provide will be used only within the context of the researcher's Ph.D. thesis work and that my confidentiality and the confidentiality of my organization will be respected through the use of an alias which will not be linked to my name or to my organization's name, electronically or otherwise. The interviews will be recorded; the transcripts produced from the recordings will only be available to the researchers conducting this study and will be stored in a locked cabinet and destroyed after five years. If I decide to participate in only one of the two interviews, the data collected at that time will be used in the study.

I understand that since these activities deal with my professional opinion, I may feel uncomfortable with some of the questions asked during the interviews. I have received assurances from the researchers that every effort will be made to minimize these risks: I am free to withdraw from the research project at any time; to refuse to go on; or to refuse to answer particular questions.

Any information about my rights as a participant in this study may be addressed to the Protocol Officer for Ethics in Research, 550 Cumberland Street, Room 159, (613) 562-5600 extension 1787 or ethics@uottawa.ca. Questions about the conduct of the study can be addressed to J.B. Cousins, Ph.D., at the Faculty of Education, (613) 562-5800 extension 4036 or bcousins@uottawa.ca.

I may keep a copy of this letter and refer to it in any communications with the researcher, the Ethics Officer, or other staff members of the University of Ottawa. I understand that by signing below, I am stating my consent to participate in the study.

Signature of participant: _____ Date: _____

Signature of researcher: _____ Date: _____

____ I wish to receive a summary of the research results and copies of publications based on this project.

APPENDIX C

Phase 1 - Interview Protocol

Thank you very much once again for agreeing to participate in this interview. As I mentioned in my email, I will be recording our conversation to make sure that I don't miss anything while we talk. Is this still ok with you?

If yes: Before we start, I just want to reiterate that everything you say in this interview will be treated as confidential, and that you will not be identified by name or other means in the research report.

Any questions or comments before we start?

1. The reason why I asked you to participate in this study is because of your broad knowledge of the evaluation function in the federal government. I am interested in identifying departments or agencies at different levels of evaluation capacity in order to identify the factors that promote or hinder capacity development in these organizations. So, first of all, I'd like to ask you about your own definition of evaluation capacity. What do these words mean for you?
2. What would you look for in an organization in order to assess its evaluation capacity? *(prompt: In your opinion, what are the key elements or dimensions of evaluation capacity? How would you define each of these elements or dimensions?)*
3. Can you give me some examples of Canadian federal government departments or agencies that you think reflect a high or exemplary level of evaluation capacity? Why? What specific characteristics do these organizations have that make them highly capable in terms of evaluation practice?
4. Can you give me some examples of Canadian federal government departments or agencies that you think reflect a lower level of evaluation capacity? In what sense is their capacity lower than other organizations?
5. What concrete steps have been taken by organizations that you're familiar with to build their evaluation capacity? In your opinion, is this something that is important to government organizations? Has it worked? Why/why not?
6. Finally, do you know of any organization that has clearly benefited from building its evaluation capacity, government or otherwise? What are these benefits?

Did you want to add anything else on evaluation capacity in the federal government or on evaluation capacity building?

Thanks very much for participating.

I will be using your responses and those of other interviewees to develop a framework that describes the key dimensions of evaluation capacity in federal government organizations. This framework will be in matrix form and will highlight the essential dimensions of evaluation capacity as well as the manageable steps required to move from typical, or low, capacity to high, or ideal, capacity.

I would appreciate your thoughts on this framework once it is drafted. Are you still willing to be interviewed a second time? The next interview will focus on the framework itself, its clarity, wording, and contents.

APPENDIX D

Phase 2 - Interview Protocol

Thank you very much once again for agreeing to participate in this second interview. Once again, I will be recording our conversation to make sure that I don't miss anything while we talk. Is this still ok with you?

If yes: Before we start, I just want to reiterate that everything you say in this interview will be treated as confidential, and that you will not be identified by name or other means in the research report.

Any questions or comments before we start?

In our last meeting, I asked you about the key elements or dimensions of evaluation capacity. I used your comments and those of other participants, as well as a review of the literature, to develop a draft framework that describes evaluation capacity in government organizations. I sent you this framework via email.

1. Did you have a chance to look at the framework? (*If no: Let's take a look at it together right now.*)
2. First, I'd like your general impression of the dimensions included in the framework. (*Prompt: Does the framework include the most important dimensions of evaluation capacity? Is anything missing? Are any of the dimensions irrelevant?*)
3. Is the wording used for the different column headings clear? Does it adequately reflect each of the dimensions of evaluation capacity?
4. Next, I'd like us to go through each of the dimensions and the steps required to move from a typical level of performance to an ideal level of performance. I'd like you to tell me if you think the steps are manageable, and if they are clearly outlined in the framework. (*Prompt: Go through each step for each dimension.*)
5. Do you think that this framework could be used by organizations interested in self-assessing their level of capacity? Why or why not?
6. Finally, the next phase of my study involves testing out a revised draft of this framework in federal government organizations. Can you recommend organizations that would have a typical level of evaluation capacity and others that have an ideal level of evaluation capacity?

Did you want to add anything else on evaluation capacity in the federal government or on evaluation capacity building?

APPENDIX E

Phase 3 - Introductory Email

Dear _____,

I am currently conducting a research project on evaluation capacity in the Canadian federal government as part of my Ph.D. thesis in measurement and evaluation at the Faculty of Education, University of Ottawa. My study involves the development and validation of a framework that describes evaluation capacity in federal government departments and agencies. I have developed such a framework and am now looking to validate its contents in a number of federal government departments and agencies. As Head of Evaluation at (*insert name of department here*), I am seeking your cooperation in this process through an interview based on your use of the framework.

The interview should take approximately one hour and will be held in person, at your place of business or other location of your choice. Your participation will involve using the framework by following its accompanying directions and providing feedback on your experience in the interview. With your permission, I would like to record the interview 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.

If you are willing to participate in this study, please reply to this email indicating when you are available over the next few weeks or call _____ to speak with me directly. I would be happy to answer any questions on the research project or the interview prior to our first meeting, or when we meet. You may also address questions or concerns to Professor J. Bradley Cousins at the University of Ottawa, who is supervising this thesis (613.562.5800 ext. 4036 or bcousins@uottawa.ca).

Thank you very much for taking the time to consider this request. I look forward to hearing your insights on this interesting topic.

Sincerely,

Isabelle Bourgeois
Ph.D. Candidate, Faculty of Education
University of Ottawa

APPENDIX F

Phase 3 - Letter of Informed Consent

I am currently conducting a research project on evaluation capacity in the Canadian federal government as part of my Ph.D. thesis in measurement and evaluation at the Faculty of Education, University of Ottawa. My study involves the development and validation of a framework that describes *evaluation capacity in Canadian federal government departments and agencies*. I have developed such a framework and am now looking to validate its contents in a number of federal government departments and agencies. Your organization has agreed to take part in this research and I would like to invite you to participate in this final phase of my study.

Your participation would involve the application of an evaluation capacity growth profile framework to your organization and one follow up in-person interview with myself. The interview would take place at a time deemed suitable for you. It is expected that application of the framework will require approximately one hour and the interview will last for about an hour.

The information you provide would be used only within the context of my Ph.D. thesis work and your confidentiality will be respected by my withholding your identity and that of your organization in any published reports. With your permission the interview will be audio-taped; the transcripts produced from the recordings will be stored in a locked cabinet and will be destroyed after five years.

These activities would deal with your professional opinion about your organization. You would be free to decline answers to specific questions or to withdraw from the research project at any time.

Enquiries about the study or your rights as a participant may be addressed to the **Protocol Officer for Ethics in Research**, 550 Cumberland Street, Room 159, (613) 562-5600 extension 1787 or ethics@uottawa.ca, or to my thesis supervisor, Professor Brad Cousins, Ph.D., at the Faculty of Education, (613) 562-5800 extension 4036 or bcousins@uottawa.ca.

Thank you for considering this request. Your signature below would indicate your willingness to participate in the study.

Signature of participant: _____ Date: _____

Signature of researcher: _____ Date: _____

____ I wish to receive a summary of the research results and copies of publications based on this project.

APPENDIX G

Phase 3 - Interview Protocol

Thank you very much once again for agreeing to participate in this interview. Once again, I will be recording our conversation to make sure that I don't miss anything while we talk. Is this still ok with you?

If yes: Before we start, I just want to reiterate that everything you say in this interview will be treated as confidential, and that you will not be identified by name or other means in the research report.

Any questions or comments before we start?

At the beginning of this research project, I asked a number of evaluation experts about the key elements or dimensions of evaluation capacity. I used their comments, as well as a review of the literature, to develop a draft framework that defines evaluation capacity in government organizations. I sent you this framework two weeks ago via email.

1. Did you have a chance to look at the framework? *(If no: Let's take a look at it together right now – skip to question 3).*
2. Did you use the framework in your discussions with colleagues or on your own?
3. First, I'd like your general impression of the dimensions included in the framework. *(Prompt: Does the framework include the most important dimensions of evaluation capacity? Is anything missing? Are any of the dimensions irrelevant?)*
4. Is the wording used for the different column headings clear? Does it adequately reflect each of the dimensions of evaluation capacity?
5. Next, I'd like us to go through each of the dimensions and the steps required to move from a typical level of performance to an ideal level of performance. I'd like you to tell me if you think the steps are manageable, and if they are clearly outlined in the framework. *(Prompt: Go through each step for each dimension).*
6. Can you tell me about your organization's evaluation capacity, based on the framework? *(Prompt: Use the framework to assess your organization's evaluation capacity).* How difficult was it to use the framework? Do you think that you would be able to use it with your colleagues to discuss evaluation capacity?

Thanks very much for participating. I will send you a final draft of the framework once it's ready, in case you'd like to use it to discuss evaluation capacity with your colleagues. Any final comments on this process?

APPENDIX H

Phase 3 - Letter of Organizational Consent

I am currently conducting a research project on evaluation capacity in the Canadian federal government as part of my Ph.D. thesis in measurement and evaluation at the Faculty of Education, University of Ottawa. My study involves the development and validation of a framework that describes *evaluation capacity in Canadian federal government departments and agencies*. I have developed such a framework and am now looking to validate its contents in a number of federal government organizations. The purpose of this letter is to invite your organization to participate in this final phase of my study.

The participation of your organization would imply application of the framework by evaluation staff and other decision makers within your department or agency as well as one interview conducted with each of the individuals involved in using the framework. It is expected that three individuals from your organization would participate in the project (evaluation head, senior decision maker, evaluation staff person). The application of the framework may require one hour or so, and the subsequent interview conducted with each participant would take approximately an hour, for a total of two hours' involvement per person.

The potential benefits of participating in this study include gaining a better understanding of your organization's evaluation capacity as well as the identification of the manageable steps required to increase capacity along a number of defined dimensions.

Please rest assured that individual participation to this project will be entirely voluntary, and that participant and organizational confidentiality will be ensured through the use of aliases that will not be linked to individual or organizational names, electronically or otherwise. Individual participants will be asked to read and sign consent forms indicating how the data will be stored and used. Please find attached a copy of this consent form.

This research has been approved by the University of Ottawa Research Ethics Board. Enquiries may be addressed to the **Protocol Officer for Ethics in Research**, 550 Cumberland Street, Room 159, (613) 562-5600 extension 1787 or ethics@uottawa.ca, or to my thesis supervisor, Brad Cousins, Ph.D., at the Faculty of Education, (613) 562-5800 extension 4036 or bcousins@uottawa.ca.

Thank you for considering this request. I truly hope that your organization would be willing and able to participate in the study. Please indicate your agreement by returning a signed copy of this letter to me at the address provided. If you have any questions, you may contact me directly (see coordinates below) or Professor Cousins. We would be happy to provide further information if necessary.

Sincerely,

Isabelle Bourgeois
Ph.D. Candidate, Faculty of Education, University of Ottawa

My organization agrees to participate in the study.

Signature: _____ Date: _____