PAINTING A PORTRAIT OF ORGANIZATIONAL EVALUATION CAPACITY IN THE CANADIAN ART MUSEUM SECTOR

by

Agnieszka Chalas

A dissertation submitted to the Faculty of Education
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
(April, 2019)

Copyright ©Agnieszka Chalas, 2019
Abstract

Not only is research on program evaluation practice and capacity in art museums largely absent, but also the actual dimensions of evaluation capacity as they could be observed in these unique professional settings have heretofore neither been conceptualized or defined based on empirical data. This study sought to (a) develop a framework that conceptualized what evaluation capacity might look like in art museums and (b) examine how such capacity manifested itself across the framework’s various dimensions both sector-wide and in those Canadian art museums that were most active in conducting a wide range of research and evaluation studies. A two-phase multiple method qualitative research design was used to address the purposes of this research. Phase One involved conducting an interview study to establish an initial knowledge base on Canadian art museum educators’ program evaluation practices and capacities and test the degree to which the initial conceptual framework that was developed to guide this study could be considered an accurate and complete description of evaluation capacity in the Canadian art museum context. Phase Two involved conducting qualitative case studies of two art museums that, based on the interview findings, were identified as operating at the highest level of capacity for evaluation in the country. The study provided the evidence necessary to finalize the initial conceptual framework and concluded that evaluation capacity in Canadian art museums could be described through six central sub-divided dimensions. The study results likewise both painted a portrait of moderate capacity for evaluation across the sector (with smaller pockets of high capacity) and shed empirical light on the phenomenon of developed capacity in selected Canadian art museums. In, demarcating the dimensions that comprise evaluation capacity in art museums, this research makes a significant theoretical contribution to the evaluation literature. Several key recommendations that outline what could be done to
strengthen the evaluation capacity of art museums in Canada, meanwhile, represent the main practical implication of this study. These recommendations are likely to be useful not only to the growing number of art museums seeking to integrate evaluation into their organizational cultures but also to several other sectors and organizational types.
Acknowledgements

I want to both acknowledge and thank those individuals that were instrumental to the development of this dissertation beginning with the art museum practitioners that graciously agreed to participate in this study. This dissertation would not have been possible without their openness and genuine interest in program evaluation and other forms of visitor studies inquiry. My deepest gratitude goes to my supervisor, Dr. Chris DeLuca, for guiding and advising me during these past few years. I felt extremely lucky to have a supervisor who always made himself available when I needed him and genuinely cared about both my academic and professional success. To the members of my dissertation committee, Drs. Rena Upitis and Jean King, thank you for your support and valuable suggestions at each step of the research process. Your varied expertise helped to strengthen this dissertation in so many ways. To the chair and members of my dissertation defense examining committee, Drs. Joan Schwartz, Lyn Shulha, Isabelle Bourgeois, and Sascha Priewe, thank you for pushing my thinking further with your insightful questions and helpful suggestions. I would also like to acknowledge both the Social Sciences and Humanities Research Council of Canada and Queen’s University for generously funding my work. A heartfelt thank you goes to my husband and thought partner, Michael, who was a great source of support and encouragement throughout this process—I could not have done it without you! Finally, I want to dedicate this dissertation to my beloved son, Henry, for whom I undertook this academic journey in the first place.
# Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgements ............................................................................................................... iv
List of Figures ......................................................................................................................... xii
List of Tables ......................................................................................................................... xiii
Chapter 1 Introduction ......................................................................................................... 1
  Purpose ................................................................................................................................. 4
  Rationale ............................................................................................................................... 5
  Defining Key Terminology ................................................................................................. 7
  The Art Museum Context .................................................................................................... 10
    Educational programming in art museums ...................................................................... 10
    The role of program evaluation within museum visitor studies ..................................... 12
    The (im)possibility of evaluating the outcomes of art museum education ..................... 16
    The art museum as a unique context ................................................................................ 18
  Researcher Background and Reflexivity .......................................................................... 19
  Dissertation Overview ....................................................................................................... 20
Chapter 2 Literature Review and Conceptual Framework ................................................. 21
  The Profile Framework of Organizational Evaluation Capacity ..................................... 24
  The Conceptual Framework of the Capacity to Do and Use Evaluation ......................... 32
  The Conceptual Model for the Measurement of EC ........................................................ 37
  The Synthesis Model of Evaluation Capacity .................................................................... 40
  Comparison of Frameworks .............................................................................................. 43
    Overlap of constructs ........................................................................................................ 43
Internal impetus .................................................................................................................. 99
External pressure ................................................................................................................ 100
Resources and Supports .................................................................................................... 101
Time and human resources ................................................................................................. 101
Financial resources and other supports ............................................................................. 102
Capacity to Do and Use Program Evaluation ..................................................................... 104
Knowledge and skills ........................................................................................................ 104
Use of evaluation findings .................................................................................................. 107
The impact of evaluation processes .................................................................................... 112
Program Evaluation Activities ........................................................................................... 114
Type of evaluation engaged in based on evaluation purpose .............................................. 114
What gets evaluated ........................................................................................................... 116
Data collection, analysis, interpretation, and reporting ...................................................... 117
Type of information collected ............................................................................................ 119
Who was involved in the evaluation process .................................................................... 120
Satisfaction with the frequency and quality of program evaluation activities ............... 122
Evaluation Capacity Building ............................................................................................ 122
Practice-Based Challenges and Limitations to Capacity ...................................................... 125
Practitioners’ Self-Reported Evaluation Needs .................................................................. 127
Reactivity and Opportunities for the Future ........................................................................ 130
Chapter Summary ............................................................................................................. 132
Chapter 5 Phase Two Findings ........................................................................................ 134
Case Portrait 1: The Montreal Museum of Fine Arts ......................................................... 134
Ongoing pan-institutional engagement with a wide range of research and evaluation activities........................................................................................................ 180
The presence of an evaluation-minded culture........................................................................................................ 181
Dedicated resources and supports........................................................................................................ 181
Concluding comments ........................................................................................................ 182

Chapter 6 Summary, Discussion, and Implications ........................................................................ 184

Dissertation Purpose, Questions, and Study Design ................................................................. 184

Discussion of Key Findings by Research Question ...................................................................... 187
Key findings pertaining to research question number one ......................................................... 187
Key findings pertaining to research question number two ........................................................... 193
Key findings pertaining to research question number three ....................................................... 199

Implications of Key Findings ........................................................................................................... 203
Implications for the framework ...................................................................................................... 203
Implications for practice ................................................................................................................ 207
Implications for research ............................................................................................................... 213

Conclusion ........................................................................................................................................... 216

References ........................................................................................................................................... 217

Appendix A Letter of Ethics Clearance from Queen’s University .............................................. 240
Appendix B Sample Email Scripts ................................................................................................. 241
Appendix C Example Telephone Scripts ...................................................................................... 243
Appendix D Phase One Combined Letter of Information and Consent Form ......................... 246
Appendix E Phase One Interview Protocol .................................................................................. 248
Appendix F Start List of Codes ...................................................................................................... 252
Appendix G Sample Matrix Coding Charts................................................................. 254
Appendix H Phase Two Combined Letter of Information and Consent Form............... 255
Appendix I Phase Two Interview Question Matrix....................................................... 262
Appendix J Final List of Codes Applied During Phase One ........................................ 265
Appendix K Final List of Codes Applied During Phase Two ........................................ 270
Appendix L Sample of Documents and Documentary Data Analyzed Per Museum During Phase Two ........................................................................................................................................... 278
List of Figures

Figure 1. Museum visitor studies inquiry domains .......................................................... 13

Figure 2. The exhibition program development and evaluation process.......................... 15

Figure 3. The conceptual framework of evaluation capacity in art museums.................... 53

Figure 4. The location of EC diagnosis within the ECB process .................................... 63

Figure 3. The conceptual framework of evaluation capacity in art museums (revised)......... 205
List of Tables

Table 1. Evaluation capacity studies reviewed ......................................................... 23
Table 2. Canadian art museums sampled ................................................................. 67
Table 3. Overview of participants and data sources per museum ......................... 70
Table 4. List of case study informants per museum ............................................... 81
Table 5. Overview of data collection methods used .............................................. 118
Table 6. Overview of select ongoing research and evaluation projects ................. 149
Chapter 1

Introduction

Museums are no longer their own excuse for being. As the resources they require have become greater and greater, so, too, have the expectations of those called upon to provide those resources. What is demanded today is that organizations perform, deliver, and demonstrate their effectiveness. (Weil, 2003, p. 53)

During a Skype interview in early 2015, Judy Koke, former Chief of Public Programming and Learning at the Art Gallery of Ontario (AGO), told me that evaluation of the gallery’s educational programs is “certainly something we’re trying to do and we have done. It’s not like we don’t do it. We just don’t do it systematically, and we don’t do it as well as we should.” Ms. Koke’s colleague, Keri Ryan, Director of Interpretation and Visitor Research, echoed this sentiment, explaining, “There’s a real appetite for making data-driven decisions here now, so there’s no shortage of desire for program evaluation. It’s just that we haven’t quite caught up in terms of how we resource it.” Their admissions struck me as confirmatory, rather than revelatory.

For three years, I held the role of Director of Public Programs and Education at the Kitchener-Waterloo Art Gallery (KWAG) where I held broad responsibilities, including conducting internal program evaluations (Chalas, 2015, 2017; Chalas & Reid, 2017). Like many of my colleagues, I received no formal training in evaluation as part of my professional preparation; I was, in many ways, an accidental evaluator (King & Volkov, 2005). I can’t help but recall one Family Sunday event at KWAG where, while greeting visitors as they signed in, I collected their postal codes and email addresses and tracked attendance figures using a hand-held
tally counter. Click. Later, I came to realize that while such measures may have provided me with valuable information about how many people were visiting and their demographics, these data did not describe whether the popular program was meeting its learning-related objectives or how participants benefitted from their experiences. Neither, for that matter, did the parent satisfaction surveys I administered on an annual basis in fulfillment of my granting requirements. I wanted to do a better job of program evaluation at the gallery so that I could better meet the needs of diverse participants. I envisioned myself employing a wide range of evaluation methods, both qualitative and quantitative, to gather convincing evidence of the outcomes achieved by the educational programs\(^1\) we offered. I saw myself collaborating with participants in evaluative efforts—especially in the case of programs in which they had been actively involved. I imagined using the evaluation process to actively build internal evaluation capacity among staff members so that we could make evidence-informed decisions about how to improve our programs together. How might KWAG broaden its evaluation agenda to match this picture I wondered? What were other art museums doing to evaluate the consequences of their educational programs?

With these questions in mind, I reached out to the AGO as a first step towards examining the overall state of program evaluation in the Canadian art museum sector. My conversations with Ms. Koke and Ms. Ryan revealed both the barriers that the AGO, one of North America’s largest museums, has faced on its journey toward evaluating programmatic impact and the strategies the gallery has employed to enhance its internal evaluation efforts (Chalas, 2016). They also reassured me that I was not alone in thinking that it is time for art museums to start

\(^1\) The term *educational program* is used to refer to a broad range of programs offered by a museum’s education/public programming/learning department from school programs to adult programs to special projects etc.
evaluating their educational and public programs and to evaluate them well, particularly in today’s accountability climate. As Judy Koke noted:

If museums are to survive long term, they need to be able to articulate why they matter in our communities…and I think visitor research and evaluation is the path to getting there. It’s the only way we can tell a story about the impact that we’re having so it’s really important for us to be doing more program evaluations.

Perhaps the most important takeaway from my AGO interviews, however, is how they exposed the critical need for research on program evaluation in this field. What little research exists in this area focuses on evaluation practices across a variety of different museum types—zoos, aquaria, historical societies, etc. (e.g., Reussner, 2004; Smithsonian Institution, 2004). Thus, little is known about what art museums\(^2\) specifically are doing to evaluate their educational programs, what their capacities are to do so, what factors influence their capacities, what efforts they engage in to reflect on and build these capacities, what their evaluation resource and support needs may be, or the challenges they face when responding to funders’ increased calls for accountability. Moreover, the actual dimensions of evaluation capacity (EC) as they could be observed in art museums have not yet been identified in the evaluation literature. Recognizing this, numerous scholars in the museum community have called for further investigations into program evaluation within this sector (e.g., Hein, 2005; Parsons, 2004; Sheppard, 2000, 2010). Still others working in evaluation have argued that more research is

\(^2\) In British and Commonwealth usage (i.e., English Canada), the term gallery implies a public gallery while the term museum refers to non-fine art institutions. In the United States and Quebec, on the other hand, a public gallery is described as an art museum. This is also the case in the museum literature. In keeping with this literature, I am using the term art museum to refer to public galleries in both Anglophone (i.e., the Art Gallery of Ontario) and Francophone (i.e., the Montreal Museum of Fine Arts) Canada.
needed that advances the field’s understandings of EC across a variety of sectors and organizational types (e.g., Cousins & Bourgeois, 2014; Taylor-Ritzler et al., 2013).

**Purpose**

My doctoral research addresses the foregoing gaps in the literature through a two-phase qualitative study that sought to both establish an initial knowledge base on program evaluation in Canadian art museums and contribute to the existing EC literature base by deepening the field’s understandings of organizational capacity for evaluation across a yet unexplored sector—Canadian art museums. Specifically, the purposes of this study were to

(a) identify and describe the key dimensions that comprise EC in Canadian art museums by developing and empirically validating a conceptual framework of EC for use in the art museum sector,

(b) establish a baseline of practitioners’ understandings of program evaluation (i.e., practices, capacities, building efforts, influencing factors, needs, and challenges), and

(c) obtain an in-depth understanding of what developed EC looks like.

Three central research questions guided this study:

1. What are the key dimensions that characterize evaluation capacity in the Canadian art museum sector?

2. How is program evaluation capacity manifested across Canada’s largest publicly owned art museums, and to what extent does this capacity vary from museum to museum?

3. How is overall organizational evaluation capacity manifested in those Canadian art museums most active in evaluating their educational programs, and which factors and

---

3 The term *developed* capacity refers to the highest level of EC reported by an art museum. It does not necessarily connote *exemplary* capacity nor is it meant to equate to the highest level of EC reported in another sector (i.e., government).
conditions do practitioners report having enabled their capacities to develop to their current levels?

**Rationale**

There are two central arguments supporting the importance of this study. First, as education has taken a more prominent role in museums (Hein, 1998; Weil, 2002), governments and external funding agencies have increased demands for evaluations of publicly-funded programs in the name of accountability (Gorman, 2007; Korn, 2007). As a result, it is no longer enough to simply claim that museums are inherently educational. Accountability-driven program evaluation is perhaps more pervasive in the United States where funding agencies (e.g., the National Endowment for the Arts, the Institute for Museum and Library Services, the National Science Foundation) now routinely require museums to measure the extent to which their programs achieve their intended outcomes although evidence suggests that funders’ demands for evaluation information are also on the rise in Canada (Hall, Phillips, Meillat, & Pickering, 2003). The Ontario Trillium Foundation is a case in point. The granting agency’s *Grow Grants* (e.g., *Inspired People*, Ontario Trillium Foundation, 2019) require prospective recipients to have an evaluation plan in place at the time of application and direct between 2% and 10% of the total grant value towards the comprehensive evaluation of programmatic impacts.

While such funding pressures have led to a rise in the frequency of program evaluations in art museum education over the past thirty years (Adams, 2012), they have also raised new challenges for these types of museums. Foremost among these is the fact that while some larger art museums can afford to hire in-house evaluators or external consultants to provide evaluation services (Falk 2000; Luke & Ancelet, 2014), the task of evaluation in small to mid-sized institutions falls to education staff with little or no training in evaluation (Chen-Cooper, 2007;
Ebitz, 2005) who are nonetheless required to demonstrate measurable outcomes of programs. Beleaguered by other responsibilities, many such practitioners would prefer to devote their time and limited resources to program development and delivery over evaluation. Many more either hold negative attitudes towards or are uncomfortable with evaluation, erroneously associating all program evaluation with positivism despite the rich and varied methodological options available (Hooper-Greenhill, 2004).

Moving into the future of art museum education, it is unlikely that the demand for accountability-driven evaluation will abate. Increasing museum practitioners’ understandings of and building their capacities around evaluation are therefore of paramount importance to art museums. Indeed, the field has seen a developing trend toward this end with a growing number of museums engaging in institution-wide evaluation capacity building (ECB) (Luke & Ancelet, 2014). Before art museums can effectively initiate such efforts, however, they must first diagnose an art museum’s capacity for evaluation. Cousins, Bourgeois, and Associates (2014) explain, “It is imperative to learn more about what capacity looks like to better inform strategies, practices, processes, and the like, designed to improve it” (p. 26). Consequently, through the development and validation of a conceptual framework uniquely suited to describing evaluation practice and diagnosing EC in art museums, this study will increase the ability of scholars, ECB developers, and museum practitioners to measure EC in art museums and identify appropriate ECB activities aimed at developing that capacity.

A second related argument for the importance of this research concerns the lack of relevant examples of evaluation practice for art museum professionals to turn to for guidance. For instance, even those art museum professionals who recognize the value of evaluation and are evaluating their educational programs to the extent that their capacities allow are challenged by a
lack of published research on evaluation in this sector as well as a lack of available evaluation reports that might serve as examples for their own work (Adams, 2012; Luke & Ancelet, 2014). Taking an in-depth look at those art museums that are most active in evaluating their educational programs may therefore be key to developing a better understanding of what good evaluation practice looks like than is presently the case and to sharing such practice with other art museums interested in increasing the quality and frequency of their evaluations. In this way, the results of this study will contribute knowledge aimed at improving evaluation practice and increasing capacity for evaluation across the Canadian art museum sector.

**Defining Key Terminology**

Within this study, defining *program evaluation* is of crucial importance. Encompassing a wide variety of approaches, program evaluation is a professional practice dedicated to making judgments about programs. For the purpose of this paper the following definition as presented by Patton (2008) is adopted: “Program evaluation is the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding” (p.39). Consequently, throughout my research the term *program evaluation* will be used to refer to evaluating the intended and unintended consequences of participation in formal programs that *education departments*\(^4\) in Canadian art museums offer. Excluded from such a definition is both the evaluation of an art museum’s *exhibition programs* and the evaluation of what people learn from engaging with *exhibits* during causal drop-in museum visits.

\(^4\) The term *education department* is used to refer to the division in an art museum responsible for developing and implementing programmatic offerings for a wide range of audiences irrespective of what that division is titled in practice (i.e., *Public Programming and Learning* at the Art Gallery of Ontario or *Education* at the Winnipeg Art Gallery).
While often blurred in the evaluation literature, it is also critical to establish a distinction between evaluation capacity (EC) and evaluation capacity building (ECB) for this research and to consider the variation in definitions used within the evaluation field today. One oft cited and widely accepted definition of ECB is that of Compton, Baizerman, and Stockdill (2002). The authors define ECB as:

A context-dependent, intentional action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/programs/sites. (p. 109)

Intentionally developing an organization’s evaluation capacity has been observed to enable the institutionalization of evaluation practice (i.e., making evaluation an ongoing activity in the organization).

Considerably less consensus surrounds the definition of EC. Bourgeois and Cousins (2008), for instance, define the construct as “an organization’s visible, enacted evaluation practices and processes” (p. 128). Cheng and King (2017) define EC simply as “an outcome of ECB” (p. 522). Bourgeois, Whynot, and Thériault (2015) offer a more expanded definition when they refer to EC as “the competencies, and structures required to conduct high quality evaluation studies (capacity to do), as well as the organization’s ability to integrate evaluation findings into its decision making process (capacity to use)” (p. 47). Patin (2013) suggests that evaluation capacity be thought of as “the necessary resources (e.g., funds, knowledge, skills, time) essential to implementing and maintaining quality program evaluation in nonprofit organizations” (p. 10) while according to Nielsen, Lemire, and Skov (2011), EC is “an organization’s ability to bring about, align and sustain its objectives, structure, processes, culture, human capital and
technology to produce evaluative knowledge that informs on-going practices and decision making to improve organizational effectiveness” (p. 338).

Each of these definitions nuances different aspects or dimensions of EC and emphasizes differing understandings of the nature and purposes of both evaluation and ECB to varying degrees. For example, Nielsen, Lemire, and Skov’s (2011) definition reasons that an organization’s EC is driven by four dimensions reflecting evaluations’ demand and supply sides, situates evaluation practice within larger processes of evaluative knowledge production, and positions evaluation as an organizational management tool. That numerous definitions of EC exist in the literature is indicative of the fact that EC is still a relatively new concept in the field of evaluation, and, as such, ambiguity remains over its precise meaning. As Nielsen, Lemire, and Skov (2011) point out, the absence of a general definition of EC becomes even more apparent when the construct is further parsed out into its constituent dimensions. Despite the differing definitions and conceptualizations, however, there is consensus among scholars that EC is a complex multidimensional construct. Chapter 2 discusses, in greater detail, the overlaps and differences between the dimensions and sub-dimensions contained in the EC frameworks published to date (Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011; Taylor-Ritzler et al., 2013).

Both the terms EC and ECB have been applied at varying organizational levels. Thus, based on Bourgeois et al.’s (2015) and Compton et al.’s (2002) thinking, I define EC as a museum’s, museum department’s, or individual museum practitioner’s actual or potential ability to conduct program evaluations and make use of evaluation findings and processes and ECB as the intentional process of developing this ability to influence both program and organizational effectiveness. Albeit closely intertwined streams of inquiry, EC research focuses on asking
questions such as: “What are the actual dimensions of evaluation capacity in this context?” and “How can evaluation capacity be measured?” Key questions for ECB research on the other hand include: “How can evaluation capacity be enhanced? and “How can evaluation capacity building efforts be evaluated?” Diagnosing an organization, group, or individual’s capacity for evaluation, therefore, marks an important first step in any ECB effort, regardless of the organizational level at which it is aimed.

The Art Museum Context

Educational programming in art museums. Towards the later part of the 20th century, museums shifted their attention from objects towards their visitors and gained increased recognition as educational institutions (Buffington, 2007; Williams, 2007; Willumson, 2007). Today, education departments are ubiquitous in art museums and offer an expansive schedule of programs for a diverse range of audiences. Educational programs may complement exhibitions or be offered independently of them. They may be free, or they may charge registration fees.

Programs for school groups can range from traditional guided tours that engage an ever-changing audience to multi-part programs that engage a constant audience and unfold outside the walls of the museum. The Solomon R. Guggenheim Museum’s Learning Through Art program is an example of the latter. This sustained school-based residency sees teaching artists and classroom educators collaborate on integrated art projects that are supported by visits to the museum (Guggenheim, 2019).

Programs aimed at youth include everything from teen nights to apprenticeship programs that offer students on-the-job work experience (e.g., the New Museum’s Teen Apprentice Program, 2019) to youth-led programs such as the Art Gallery of Ontario’s Youth Council—a
year-long program that sees its members initiate both on- and off-site programming for people between the ages of 14 and 24 (AGO, 2019a).

Teacher-oriented museum programs can vary from one-time art-making workshops to week-long credited courses (e.g., the Brooklyn Museum’s Teacher Institute, 2019) to year-long paid professional development programs such as the Whitney Museum of American Art’s Teacher Exchange (Whitney, 2019).

Adult programming can encompass artist and curator talks, studio classes, 19+ networking events (e.g., the Art Gallery of Nova Scotia’s ArtParty, AGNS, 2019), and programs such as the Museum of Contemporary Art Denver’s popular Mixed Taste tag-team lectures on unrelated topics (e.g., porcini mushrooms and the American ideal) (MCA Denver, 2019).

While children’s, family, partnership, and community outreach programs are also proliferating across the sector, art museums are increasingly offering programs aimed at visitors with disabilities. These include touch tours for the blind or partially sighted, sign language tours for the deaf or hard of hearing, customized programs for students on the autism spectrum, and specialized programs for individuals with Alzheimer's disease like the Museum of Modern Art’s Alzheimer’s Project (MoMA, 2019). Further, the emergence of new technologies has seen art museums employ such modes of interpretive content and educational program delivery as podcasts, mobile applications, electronic guide books, interactive websites, and webinars. The National Gallery of Canada’s Distance Learning webinars are an example of one museum’s efforts to make its school programs virtually accessible to students outside of its immediate radius (National Gallery of Canada, 2019).

Participatory practices in museum education are also on the rise (Simon, 2010); more and more, art museums are inviting visitors to work alongside museum staff in program development
and delivery. For instance, the && Collective Residency was a program that engaged teens as collaborators at the Kitchener-Waterloo Art Gallery (Chalas, 2015). Over the course of several months, the student collective worked with education staff in the creation and exhibition of artwork and in the development of numerous activities, public events, online video content, and ephemera in response to the sciencefictionsciencefair exhibition.

It is clear from the above that there exists an extraordinary degree of variance in the educational programs offered by art museums (see also the art museum education programs survey conducted by Wetterlund and Sayre, 2009). The diversity of programs offered by any one museum—programs that both vary in their intended objectives and can result in a broad range of consequences—both complicate evaluation in this setting and necessitate selecting methods and approaches suitable to particular programs.

The role of program evaluation within museum visitor studies. In art museums, the evaluation of educational programs (as differentiated from exhibition programs) takes place within the broader discipline of museology known as visitor studies. Defined as “the interdisciplinary study of human experiences within informal learning environments” (VSA, visitor studies, para 1, 2019a), visitor studies practice is concerned with the systematic collection and analysis of data to inform decisions about interpretive exhibits and programs. While the history of museum visitor studies dates back to the late nineteenth century (Hein, 1998; Loomis, 1987), contemporary visitor studies comprise three distinct domains of inquiry—market research, scholarly research, and evaluation (Kelly, 2004; Korn, 2007; Reussner, 2003). Figure 1 below describes these three inquiry domains.

---

Visitor studies is a field of practice also known as visitor research and audience research. I am using visitor studies in this paper to encompass all these interchangeable terms.
In practice, the boundaries among these domains of inquiry are sometimes so blurred that it is difficult to distinguish one type of study from another (Korn, 2007). However, it is generally accepted that market researchers (e.g., evaluation and marketing consultants, museum marketing staff) conduct studies to understand actual and potential audiences’ perceptions, attitudes, demographics, leisure habits, and visiting patterns (see Randi Korn & Associates, 2011).

<table>
<thead>
<tr>
<th>Who</th>
<th>Scholarly Research</th>
<th>Market Research</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scholars in universities and museum research centres Evaluation consultants Museum practitioners</td>
<td>Evaluation and marketing consultants Marketing staff</td>
<td>Museum practitioners Internal evaluators External evaluation consultants</td>
</tr>
<tr>
<td>Focus</td>
<td>Visitor experiences, learning, interactions, etc.</td>
<td>Visitor perceptions, attitudes, demographics, leisure habits, visiting patterns, etc.</td>
<td>Visitor experiences, understandings, learning, pathways etc.</td>
</tr>
<tr>
<td>Use</td>
<td>Generate new knowledge</td>
<td>Identify market segments</td>
<td>Determine an exhibition or educational program’s effectiveness</td>
</tr>
<tr>
<td>Outputs</td>
<td>Research reports and scholarly publications</td>
<td>Marketing and audience research reports</td>
<td>Formal evaluation reports, executive summaries, oral reports on findings, outlines of observations and implications</td>
</tr>
</tbody>
</table>

Figure 1. Museum visitor studies inquiry domains

Scholars in the wider research community (e.g., universities, consulting firms, museum research centres) study visitor experiences and learning in museums. Emerging from this burgeoning base of scholarly literature is research examining the educational outcomes of both guided tours and formal museum programs (e.g., Greene, Kisida, & Bowen, 2014; Randi Korn & Associates, 2010; Roe et al., 2016). In addition, various theories of museum learning have been advanced from Hein’s (1998) theory of the constructivist museum to Housen's (1987, 2000–2001) stage theory of aesthetic development to Falk and Dierking’s (1992, 2000) contextual model of learning in museums. Also put forward in this literature are numerous pedagogical

Finally, those responsible for evaluation in museums who often include a mix of program staff, internal evaluators, or external consultants, undertake evaluation studies to measure the effectiveness of museum exhibits and determine the impact of educational programs on participants (see Adams & Cotter, 2011, and Randi Korn & Associates, 2014 for examples of both types of evaluation projects). A dominant approach to both the evaluation of exhibition programs and educational programs in museums consists of four separate phases that occur throughout program design and implementation—*front-end evaluation, formative evaluation, remedial evaluation*, and *summative evaluation* (Kelly, 2004; Korn, 2007).

*Front-end evaluations* are undertaken to gauge visitors’ interests in and understandings of the concepts behind a program or exhibition. *Formative evaluation* occurs during the development phase and involves either piloting a program or prototype-testing exhibition
components. *Remedial evaluation* is conducted immediately after the opening of an exhibition or launch of a new program for improvement purposes. The final phase of museum evaluation, *summative evaluation*, takes place at the conclusion of an exhibition or program to determine their success against stated goals and objectives (Kelly, 2004; Korn, 2007). Figure 2 below illustrates how these four phases are applied to the development and evaluation of *exhibition* programs—from initial conception through to installation and beyond—and where both museum visitors and practitioners input into this process.

<table>
<thead>
<tr>
<th>Audience Input</th>
<th>Program Development Phases</th>
<th>Institutional Input</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conceptual Development</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Front-End Evaluation</td>
<td>Revisions</td>
</tr>
<tr>
<td></td>
<td>Goals &amp; Objectives</td>
<td>Content and Interpretation Analysis</td>
</tr>
<tr>
<td></td>
<td>Design Development</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Formative Evaluation</td>
<td>Revisions</td>
</tr>
<tr>
<td></td>
<td>Final Design</td>
<td>Design and Technical Development</td>
</tr>
<tr>
<td></td>
<td>Installation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remedial Evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summative Evaluation</td>
<td>Revisions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Institutional Review</td>
</tr>
</tbody>
</table>

*Excellent Program*


**Figure 2.** The exhibition program development and evaluation process

In the current century, visitor studies practice is increasingly focused on evaluating educational programs; yet there is variability in the number and types of museums actually engaged in this work. Some museums, including art museums, still reflect visitor studies’

Across museum visitor studies’ three inquiry domains (i.e., scholarly research, market research, and evaluation), studies draw on diverse theoretical views, various communities of inquiry, and methodologies from a wide range of disciplines. Surveys, interviews (e.g., exhibition entry and exit interviews), and unobtrusive observations (e.g., timing and tracking studies) are the most frequent data collection methods employed in museums (Grack Nelson & Cohn, 2015), often in concert with other methods such as focus groups, participatory and informant-made video, and audio- or video-recording visitors’ behaviours and conversations (Borun & Korn, 1999; Diamond, 1999; Lachapelle, 1999; Leinhardt, Crowley, & Knutson, 2002; Screven 1990). An awareness of the specific types of activities undertaken in art museums is therefore crucial to an informed understanding of evaluation in this context.

The (im)possibility of evaluating the outcomes of art museum education. While there are inherent challenges to measuring learning in any setting, there is considerable discourse in the field of museum education about the challenges associated with program evaluation in museums or what Hooper-Greenhill (2004) calls the (im)possibility of measuring informal learning. It stands roughly as consensus that measuring program outcomes, especially learning outcomes, in free-choice (Falk, Donovan, & Woods, 2001) environments like art museums is fraught with unique challenges (Michalchik & Gallagher, 2010; National Research Council, 2009). The following challenges are among the most commonly articulated in the literature:

• maintaining contact with program participants who may no longer be actively involved
with a museum in order to measure long-term impacts proves difficult (Hein, 1998; National Research Council, 2009),

- inserting data collection into programs in a way that does not interrupt or change participants’ experiences or violate their expectations is easier said than done (Hein, 1998),
- making judgments about participants’ progress is challenging in the absence of baseline data on their prior knowledge and abilities (Hooper-Greenhill, 2004), and
- selectively constraining participants' experiences and exposure to educational activities (e.g., random assignment into different treatment conditions) can be both impractical and viewed as fundamentally at odds with the leisure-motivated nature of informal learning (Allen et al., 2007).

Faced with these challenges, many museums struggle to demonstrate measurable outcomes of programs. All too often the success of programs is measured based on outputs such as attendance figures, repeat visitation, and visitor satisfaction (Korn, 2008; Luke & Ancelet, 2014) even though funders are increasingly requesting outcome information⁶ (Hall, Phillips, Meillat, & Pickering, 2003) and scholars agree that such outputs, while being relatively easy for museums to collect, do little to describe the benefits of art museum education (Luke & Ancelet, 2014; Falk & Sheppard, 2006; Preskill, 2011; Worts, 2006).

An additional challenge to measuring learning in museums is the fundamental objection to defining outcomes for program participants in the first place. Hooper-Greenhill (2004) points out that while some participants may seek out museum programs where desired learning outcomes have been pre-identified, an equal number may not wish to focus quite so intently on

---

⁶ Information about the impact programs have on participants rather than information about program activities, the number of programs offered, or the number of visitors served (i.e., outputs)
specific learning achievements. For this reason, she argues that although the outcomes of learning may be anticipated by museums and possibly expected, it is inappropriate to require participants to achieve the specific outcomes that educators value. This line of argumentation has led some in the museum education community to forego prescribing formalized outcomes altogether in favor of inviting participants to define their own desired outcomes (Hooper-Greenhill, 2004; National Research Council, 2009).

Because of these challenges, little is known about the impact or outcomes of learning in museums, especially art museums (Hooper-Greenhill, 2004; Luke & Knutson, 2010). While the art museum education field remains a long way from resolving some of the practical and theoretical challenges inherent in evaluating educational programs in museums, acknowledgement of their existence provides valuable insight into where the field needs to go with respect to evaluation.

The art museum as a unique context. Art museums differ from other museum types (e.g., science, natural history, anthropological) in important ways. Science museums, for example, can be distinguished from art museums on several fronts. First and foremost, while education is just one of multiple competing agendas (others being the acquisition, care, and exhibition of artworks) in art museums, most science museums place learning at the center of what they do (Luke & Knutson, 2010). It is therefore not uncommon to see science museums employ a team-based approach to exhibition development in which educators are essential. While the art museum landscape is beginning to change in this regard (Wild Czajkowski & Hudson Hill, 2008), educators in art museums are typically called in to animate, mediate, and interpret curators’ messages once the exhibition development process is complete.
Second, while both types of museums value education, albeit to varying extents, science museum education and art museum education are quite distinct in purpose. Science museums see themselves as integral parts of a wider infrastructure of science education and focus on developing program participants’ science-specific knowledge and skills (Luke & Knutson, 2010). Art museum education, on the other hand, de-emphasizes discipline-based learning (e.g., artistic skills and concepts) while privileging soft-skills development and individual meaning-making (Meszaros, 2006).

Third, most visitor studies have been conducted in science museums as this is where centres dedicated to such practice are most often based. Thus, there exists a much smaller base of studies for art museums to draw upon. In examining evaluation capacity in any museum, it is imperative to remember that museums are not a homogenous group.

**Researcher Background and Reflexivity**

My proposed doctoral research combines my interest in program evaluation and art museum education. Having been involved with several Canadian art museums (i.e., the Surrey Art Gallery, the Vancouver Art Gallery, the Canadian Centre for Architecture, and the Kitchener-Waterloo Art Gallery) over the course of a decade-long career in museum education where, among other tasks, I was responsible for commissioning, supervising, participating in a wide range of visitor studies and program evaluations and carrying them out (e.g., Chalas, 2015, 2017; Chalas & Reid, 2017; Reid, 2012), I have an in-depth understanding of the Canadian art museum sector and the challenges inherent in evaluating educational programs in art museums. My own past evaluation struggles as an *accidental evaluator* (King & Volkov, 2005) in these contexts have also led to a pragmatic interest in (a) empowering art museum practitioners with the skills necessary to both *conduct* their own evaluations and make *use* of evaluation findings
and processes for both accountability and learning purposes and (b) sharing exemplary practices around program evaluation and capacity building in art museums. I believe that my previous experience in the sector will allow for a more informed analysis of the current state of EC across Canadian art museums.

**Dissertation Overview**

This dissertation is presented in six chapters. In the first chapter, I introduce the study’s purpose, research questions, rationale, and key terms as well as provide the context for the dissertation by way of a brief introduction to art museums. Following a review of the available literature on organizational evaluation capacity (EC), the development of an EC framework for use in the art museum sector represents the main contribution of the second chapter. The third chapter describes the two-phase multiple method qualitative research design that was used to address the purposes of this study and provides a rationale for the methodologies selected. In the fourth chapter, I report the results of my conversations with 13 art museum professionals who agreed to participate in the first phase of this research—an interview study focusing on their current capacities to evaluate their various educational and public programs. Chapter five both reports findings from two case studies of Canadian art museums that, based on the Phase One findings, were identified as operating at the highest levels of capacity for evaluation in the country and identifies the similarities and differences that arose across cases. Finally, in chapter six, I revisit the dissertation problem, purpose, questions and provide a brief overview of the study. This is followed by both a summary and discussion of the key findings from each phase of this study and their implications for theory, policy, practice, and future research. I conclude the chapter by briefly outlining the study’s significance.
Chapter 2

Literature Review and Conceptual Framework

Given the fact that evaluation capacity (EC) has only recently been conceptually distinguished from evaluation capacity building (ECB) (Bourgeois & Cousins, 2008), it is unsurprising that it has not, as numerous authors (e.g., Cousins & Bourgeois, 2014; Nacarrella et al., 2007; Nielsen et al., 2011) point out, received as much attention in the theoretical and empirical literature as ECB. Nevertheless, the evaluation field has seen a growing body of publications focused on EC itself. This literature, while still quite limited, has led to various frameworks and instruments seeking to define and measure organizations’ capacity for evaluation across several sectors. In this section, I review the available literature on organizational EC to develop a framework that conceptualizes the dimensions that comprise EC in art museums and, in so doing, address the first purpose of this study.

Specifically, my scoping review (Jesson, Matheson, & Lacey, 2011) focuses on (a) conceptualizing and measuring EC at an organizational level and (b) EC frameworks or, in one case (Cousins et al., 2008), an ECB framework (within which EC is treated as a distinct construct) that specifically set out to define and measure an organization’s capacity for evaluation. ECB frameworks oriented towards the building process (e.g., King & Volkov, 2005; Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012; Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008) are therefore excluded from this review. While ECB frameworks may share some common dimensions with EC frameworks (e.g., resources, leadership, evaluation expertise, evaluation culture, etc.), their intended purposes differ—from structuring the strategies employed in ECB (Taylor-Powell & Boyd, 2008) to summarizing the elements that could be considered in a study of ECB (Preskill & Boyle, 2008) to assisting organizations with mainstreaming
evaluation into their day-to-day operations (King & Volkov, 2005). As such, ECB frameworks are seen to complement rather than overlap with EC frameworks.

In the following section, I present each EC frameworks and associated empirical studies reviewed in turn, followed by a comparison of frameworks and a discussion of their limitations with respect to their applicability to the art museum sector. Table 1 below summarizes this literature. Next, I present the Conceptual Framework of Evaluation Capacity in Art Museums and conclude the chapter with a discussion of its unique characteristics and intended uses. First, however, I begin with a caveat. My scoping review includes those EC frameworks and studies published up to and including the end of 2016—the particular point in time at which the Conceptual Framework of Evaluation Capacity in Art Museums itself as well as all research instruments based on it were developed. I am aware that other EC frameworks and studies have been published since that time (e.g., Cheng & King, 2017; Fierro & Christie, 2017). While determining if or to what degree such frameworks may influence the dimensions that I include in the Conceptual Framework of Evaluation Capacity in Art Museums is beyond the scope of the present study, this issue I take up more thoroughly in my comments on implications for the framework in Chapter 6.
# Table 1

Evaluation capacity studies reviewed

<table>
<thead>
<tr>
<th>Reference</th>
<th>Summary</th>
<th>Framework</th>
<th>Sector</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bourgeois &amp; Cousins (2008)</td>
<td>Using findings from their analysis of 4 Canadian federal government organizations’ self-assessments of their EC, the authors examine the 4 levels of EC (described in their profile framework of EC) in relation to 4 literature-derived stages of ECB</td>
<td>The Bourgeois (2008) &amp; Bourgeois &amp; Cousins (2013) Profile Framework of Organizational Evaluation Capacity</td>
<td>Public</td>
<td>Based on a sequenced interview study</td>
</tr>
<tr>
<td>Cousins, Bourgeois, &amp; Associates (2014)</td>
<td>Multiple case study of EC in eight organizations across the education, government and nonprofit sectors</td>
<td>The Cousins et al. (2008) Conceptual Framework of the Capacity to Do and Use Evaluation</td>
<td>Public</td>
<td>Multiple case study (document analysis, interviews, focus groups)</td>
</tr>
</tbody>
</table>
Table 1 Continued

<table>
<thead>
<tr>
<th>Reference</th>
<th>Summary</th>
<th>Framework</th>
<th>Sector</th>
<th>Methods</th>
</tr>
</thead>
</table>

The Profile Framework of Organizational Evaluation Capacity (Bourgeois & Cousins, 2013)

The work of Bourgeois (2008) and her colleagues has examined organizational EC and its implications for ECB for nearly a decade. Based on the innovation profile approach developed by Leithwood and Montgomery (1987), Bourgeois and Cousins (2013) developed a multi-dimensional profile framework for describing the key behavioral manifestations of EC in the Canadian public sector. Their *Profile Framework of Organizational Evaluation Capacity* consists of six dimensions (namely *human resources, organizational resources, evaluation*...
planning and activities, evaluation literacy, integration with organizational decision making, and learning benefits). The first three of these dimensions fall into the overall category of capacity to do evaluation with the remaining three dimensions categorized under capacity to use evaluation. Each of the six dimensions in Bourgeois and Cousins’ (2013) profile framework is further broken down into a varying number of sub-dimensions described at four levels of evaluation capacity: low, developing, intermediate, and exemplary.

The first dimension, human resources, focuses on the composition of a government organization’s evaluation unit and comprises the following five sub-dimensions: staffing, evaluation logic and technical skills, communication and interpersonal skills, professional development, and leadership. More specifically, staffing refers to the balance between junior and senior evaluation positions and whether these are sufficient to manage the workload identified in a government organization’s evaluation plan. This sub-dimension also includes career progression for evaluators. The second and third sub-dimensions address the technical and interpersonal skills of government evaluators. These include skills related to the use of data collection methods and approaches, the development of evidence-informed recommendations and project management in addition to skills related to meeting clients’ (i.e., program managers) informational needs and building their trust. Professional development refers to both the internal and external professional development of evaluators based on regular assessment of their skill sets. The fifth sub-dimension, leadership, pertains to the managerial acumen of the heads of evaluation units.

The second dimension, organizational resources, is divided into three sub-dimensions: budget, ongoing data collection, and organizational infrastructure. Budget refers to the stability of the funding allocated to the evaluation unit. Ongoing data collection speaks to the degree to
which performance measurement systems\textsuperscript{7} have been integrated across an organization and the performance data collected fed into results-based management (RBM)\textsuperscript{8}. Attributed to the last sub-dimension, *organizational infrastructure*, is an organization’s governance structure, its evaluation and performance measurement policies, and the degree to which the organizational culture ensures that performance data feed into planning and reporting processes. Also included under this sub-dimension are organizational supports such as procurement services, communications, human resources, and access to information.

The third dimension, *evaluation planning and activities*, focuses on the activities undertaken by an evaluation unit and is divided into five sub-dimensions: evaluation plan, use of consultants, information sharing, external supports, and organizational linkages. The first sub-dimension assesses whether an evaluation plan (a) follows a 5-year cycle and is updated annually, (b) has been developed with senior managers, and (c) includes needs and risk assessments. This sub-dimension also includes systematic reviews of the evaluation unit itself. *Use of consultants* refers to the degree to which an organization balances evaluations conducted by internal evaluators with those conducted by external evaluation consultants. *Information sharing* refers to the degree that evaluators both gather information on new developments in policy and strategic planning and discuss evaluation projects with colleagues in their unit. Included under *external supports* is the degree to which evaluators use professional associations and published standards, are actively involved in broadening their network, and involve program managers in evaluation activities. The sub-dimension of *organizational linkages*, finally, aims to

\textsuperscript{7} Performance measurement is “the process and systems of selection, development and ongoing use of performance measures to guide decision making” (GOC, 2019, para 25).

\textsuperscript{8} Results-based management is “a comprehensive life-cycle approach to management that integrates strategy, people, resources, processes and measurements to improve decision making and drive change” (GOC, 2019, para 39).
determine the extent to which evaluators establish ties with both other key personnel (i.e., program managers and senior administrators) and organizational areas (i.e., policy development, strategic planning, and performance measurement).

The fourth dimension, evaluation literacy, is the first dimension to be included in the capacity to use evaluation category in Bourgeois and Cousins’ (2013) profile framework and consists of two sub-dimensions: involvement in evaluation, and results-management orientation. Involvement in evaluation refers to organizational staff members’ and other stakeholders’ awareness of the purpose and value of evaluation as well as their level of engagement in the evaluation process. The second sub-dimension, a results-management orientation, is seen to characterize a government organization’s culture. Such a culture is manifest when (a) senior managers prioritize managing for results, (b) staff understand organizational goals, (c) results chains are developed for programs, and (d) program managers implement performance management strategies while evaluators provide technical expertise.

The fifth dimension focuses on evaluation integration with organizational decision making and is divided into two sub-dimensions: management process, and decision support. Management process relates to the degree to which program and policy staff rely on evaluation information in their work while decision support refers to whether findings and recommendations are considered in organizational decision making.

Finally, the last dimension, learning benefits, addresses the use of evaluation information and is divided into two types of evaluation uses: instrumental/conceptual use, and process use.9 Instrumental/conceptual use describes whether findings (a) are used as a basis for action and

9 Process use refers to “individual changes in thinking and behavior, and program or organizational changes in procedures and culture, that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process” (Patton, 1997, p.90).
change and (b) impact on stakeholders’ understandings and attitudes about programs. The process use sub-dimension meanwhile aims to determine if behavioral or cognitive changes occurred by virtue of stakeholders’ proximity to and participation in evaluation activities.

The potential dimensions and sub-dimensions in Bourgeois and Cousins’ (2013) profile framework were first identified through an in-depth literature review and a series of semi-structured interviews with four experts knowledgeable about evaluation in the Canadian federal government. Once identified, a draft of the framework was then reviewed by the same four experts. To finalize the framework, a further 11 interviews were conducted with three different individuals (the head of evaluation, a senior evaluator, and a decision maker) in four federal government organizations who were asked to provide feedback on the framework and to self-assess their EC on its dimensions. These four organizations’ self-assessment results were then used to generate an EC profile for each participating organization. While, in general, government organizations’ capacity to do evaluation was higher than their capacity to use it, findings from Bourgeois and Cousins’ (2013) validation study revealed that not only could organizations be situated at varying levels of capacity (two exemplary, one intermediate, one low), but also that EC varied within each organization from one dimension to the next. Owing to the fact that government organizations may not simultaneously achieve the same level of capacity for all six dimensions, the researchers argue that EC might best be built an approach that focuses on targeting low-capacity areas and developing these areas over time.

While Bourgeois and Cousins (2013) concluded that EC in the Canadian federal government could in fact be described through the six dimensions in their profile framework, they were also quick to point out that the context within which their study was undertaken posed certain limitations to the interpretation of their findings. Most importantly, the focus on Canadian
government agencies and departments led to the identification of dimensions and sub-dimensions that were applicable to these organizations and likely to generalize well to government organizations in other jurisdictions and contexts given the “commonalities in application of measurement and evaluation systems in governance frameworks that embrace RBM and new public management” (Bourgeois & Cousins, 2013, p. 316). However, the researchers cautioned that these self-same dimensions and sub-dimensions may not be appropriate in other contexts. Further research, they argued, should therefore examine the applicability and relevance of the profile framework across sectors.

Bourgeois’ other work with colleagues extended Bourgeois and Cousins’ (2013) research, originally presented in her doctoral dissertation (Bourgeois, 2008), in several different ways. For example, Bourgeois and Cousins (2008) used the afore-mentioned four Canadian government organizations’ EC profiles (Bourgeois & Cousins, 2013) and integrated them with the four stages of ECB (traditional evaluation, awareness and experimentation, evaluation implementation, and evaluation adoption) that these authors discerned from an analysis of the ECB literature (e.g., Gilliam et al., 2003; Sanders, 2002; Torres & Preskill, 2001). In so doing, the authors delineated actionable recommendations for how these organizations could target areas in need of improvement and move through the process of ECB.

In response to requests from the Canadian federal government community, Bourgeois, Toews, Whynot, and Lamarche (2013) transformed Bourgeois and Cousins’ (2013) profile framework into a quantitative self-report instrument for measuring EC in Canadian federal government organizations. Conceived of as a practical management tool for use by heads of evaluation units, the organizational evaluation capacity self-assessment instrument calculates federal government organizations’ mean scores on items designed to measure each of the
framework’s six dimensions and provides users with a description of the EC level attained based on their results. The self-assessment instrument was validated through a three-step process that included (a) reviews of the draft instrument by three subject matter experts, (b) the collection of feedback from interested evaluators, and (c) pilot testing in five federal government organizations. The results of Bourgeois et al.’s (2013) study showed that the self-assessment instrument proved to be a valid tool for measuring EC in the Canadian federal government. However, the researchers admitted that the instrument could not readily be used by other organizations (e.g., provincial government or nonprofits) owing to its targeted focus on this specific context. Bourgeois et al. (2013) explained that while the advantage to such specificity lies in “its ease of use for these organizations and in the relevance and appropriateness of the results generated by the instrument” (p.15), they also concluded by saying that:

The self-assessment instrument cannot readily be used by other types of organizations (e.g., provincial government, nonprofits, etc.). Rather, significant adaptations will be required to both the framework of organizational evaluation capacity and the self-assessment instrument to better meet the needs of other organizations also interested in measuring their existing level of organizational evaluation capacity. (p. 15)

In their 2015 study, Bourgeois, Whynot, and Thériault did just that, adapting Bourgeois et al.’s (2013) organizational evaluation capacity self-assessment instrument to two different Canadian organizational contexts: a small nonprofit organization working in environmental protection and sustainable development, and a provincial government department. Specifically, after being reviewed by key organizational contacts and revised by the researchers based on the feedback received, the self-assessment instrument was piloted in yet another government setting before being implemented across the two participating organizations. The provincial government
organization completed the self-assessment instrument independently while the researchers completed the instrument on behalf of the nonprofit. These organizations’ self-assessment results were then compared to the results of two linked federal government agencies that participated in Bourgeois et al.’s (2013) study. In this latter case, the organization’s evaluation team completed its self-assessment during a three-hour workshop facilitated by the researchers.

Overall, Bourgeois et al.’s (2015) study revealed that all three government organizations self-assessed themselves at an intermediate level of capacity while the nonprofit organization was scored at a developing level of capacity—an unsurprising finding given the fact that the nonprofit sampled had no evaluation staff, nor any external demand for or previous experience with evaluation. The researchers’ study also highlighted the challenges inherent in creating standardized EC instruments for the purposes of inter-organizational comparison. They stated:

Generally, models and instruments such as the organizational evaluation capacity self-assessment instrument used in this study tend to be effective when used within the organizations/sectors for which they were designed, but may not prove to be applicable across different organizations or sectors, and therefore can yield findings of limited applicability. (Bourgeois et al., 2015, p. 55)

While Bourgeois et al. (2015) argue that they were able to circumvent this limitation, the researchers’ study could have been strengthened if (a) the self-assessment instrument was implemented in a similar manner across the four organizations, (b) a nonprofit with an established evaluation function was sampled, and (c) the adapted self-assessment instrument was also piloted in a nonprofit setting to ensure that it reflected the dimensions of EC appropriate to this type of organization.
The Conceptual Framework of the Capacity to Do and Use Evaluation (Cousins et al., 2008)

Cousins and his colleagues (e.g., Cousins et al., 2008; Cousins & Bourgeois, 2014) base their program of research on a conceptual framework of ECB grounded in a prior synthesis and analysis of the evaluation and the management literature (Cousins, Goh, Clark, & Lee, 2004). In Cousins et al.’s (2008) Conceptual Framework of the Capacity to Do and Use Evaluation (republished as the Conceptual Framework Depicting Organizational Capacity to Do and Use Evaluation in Cousins, Goh, Elliott, and Bourgeois [2014]), EC is conceptualized not only as a separate construct situated within ECB, but also as integral to the broader construct of organizational learning capacity. The authors’ framework is comprised of seven inter-related dimensions grouped under three main categories—antecedent conditions affecting evaluation capacity, evaluation capacity and processes, and evaluation consequences.

Included in the first category are two dimensions—sources of knowledge, skills and abilities, and organizational support structures. The dimension of sources of knowledge, skills, and abilities focuses on the development of evaluators’ competencies through both pre- and in-service training. Informal indirect ECB activities like learning-by-doing and more formal direct ECB activities such as workshops, courses, and degree or certificate programs are therefore represented in this dimension. Overlapping with sources of knowledge, skills, and abilities is a second antecedent construct originally developed as part of Goh’s (2000) conceptual framework of organizational learning capacity—organizational support structures. This dimension refers to both the level of formalization in an organization’s structure (i.e., hierarchical organizational design vs. flat) and the various training supports (including evaluation training) it invests in. Organizational support structures are also represented by the formal and informal incentive
systems government organizations implement to reward staff for the use of performance information.

The second category of Cousins et al.’s (2008) conceptual framework, *evaluation capacity and processes*, encompasses four dimensions conceptualized as being associated with the nature of organizational EC itself—*capacity to do evaluation, capacity to use evaluation, evaluative inquiry, and mediating conditions*. The *capacity to do evaluation* refers to both the knowledge and skills required to either carry out or oversee evaluation activities (e.g., data collection, analysis, reporting, etc.) and such soft skills as conflict resolution and teamwork. The *capacity to use evaluation* dimension includes process use in addition to the instrumental, conceptual, and symbolic uses of evaluation findings (Patton, 1997; 2008). *Evaluative inquiry* pertains to the nature of (i.e., internal vs. external, formative vs. summative, qualitative vs. quantitative, etc.) and extent to which specific types of evaluation activities (e.g., needs assessments, environmental scans, outcome monitoring, program evaluations, etc.) occur within organizations. The last dimension in this framework category, *mediating conditions*, refers to the factors or conditions that affect evaluation use in government organizations. These conditions include, among others, evaluation quality, evaluator credibility, timeliness, the needs of primary users, and the involvement of non-evaluator stakeholders in evaluation. Also couched under this dimension is the degree to which peer reviews or meta-evaluations support evaluations.

The remaining two framework dimensions—*organizational learning capacity* and *organizational consequences*—fall into the last category of Cousins et al.’s (2008) conceptual framework: *evaluation consequences*. *Organizational learning capacity* describes five interrelated managerial practices assumed to be essential for learning to take place in an organization. These include (a) clarity and employee support for an organization’s mission and
vision, (b) committed leadership that empowers employees, (c) an experimental organizational culture, (d) the ability of and organization to transfer knowledge and learn from failure, and (e) an emphasis on teamwork and cooperation (Goh, 2000; Goh & Richards, 1997). The organizational consequences of organizational learning capacity include, for example, shared understandings of an organization and how it operates.

Using a questionnaire based on this framework, Cousins et al. (2008) surveyed 340 internal evaluators working in both Canadian government and non-government (nonprofit, NGO, university, private) settings. To better understand how such organizations build capacity for evaluation, respondents were asked about their perceived EC in addition to their perceptions of the factors that influence ECB and evaluation use.

Descriptive survey results revealed some differences across respondent role and organizational type. Non-government respondents and those who self-identified as managers had higher scale scores for organizational learning capacity, training support, use of evaluation findings, and ECB than did respondents who self-identified as evaluators or worked in government organizations. Non-government respondents also perceived higher levels of process use and stakeholder participation and lower levels of bureaucracy. Managers reported producing reports more frequently than evaluators, and reports about program activities were produced more frequently in non-government settings. While evaluators scored slightly higher than managers on the scale score for mediating conditions, perceptions about organizational capacity to do evaluation did not differ by respondent role or type of organization.

Overall, these survey results were interpreted as painting a picture of moderate capacity to do and use evaluation while providing insight into the factors and conditions that could support greater EC in both government and non-government settings (Cousins et al., 2008). The
survey results also raised a number of considerations for future research. Specifically, survey results pointed to the need to (a) gain a greater understanding of evaluation culture within non-government settings, (b) examine in more detail the various organizational types grouped within the government and non-government categories, (c) conduct multivariate analyses to explore the relationships between and among the various framework dimensions, and (d) involve defined evaluation users in future surveys who could speak to evaluation use in their organizations. The researchers’ results also pointed to the need to explore more deeply differences between the government and voluntary (i.e., nonprofit, NGO) sectors in addition to the potential confound between sector and organizational role.

Recognizing this need, Cousins, Goh, Elliott, Aubry, and Gilbert (2014) revisited Cousins et al.’s (2008) data set in the interest of comparing government (N=160) and voluntary sector (N=89) evaluators’ survey results. Specifically, the researchers were interested in learning if and how responses between these two groups differed with respect to their self-reported understandings of EC and the antecedent conditions underlying evaluation in these organizations. Cross-tabulation analyses revealed that government respondents were considerably more likely to self-identify as evaluators than voluntary respondents (125 vs. 25)—most of whom were program managers or senior administrators with evaluation responsibilities. As such, government respondents spent more time conducting evaluations (70% vs. 25%) and reported higher levels of knowledge about evaluation theory (52% vs. 33%) and practice (65% vs. 35%) than did their voluntary counterparts. Multivariate analyses further revealed that voluntary sector respondents perceived their organizations as being less bureaucratic than government organizations, while also being more capable of learning, supportive of training, and more committed to ECB.

Although no sector differences were observed with respect to capacity to do evaluation or
for conditions mediating use, significantly higher levels of capacity to use evaluation tended to be reported in voluntary organizations. This finding was attributed to the fact that program managers in these organizations were both responsible for doing evaluations and using evaluation results and processes for decision making. The researchers explained that because few resources tend to be available for evaluation in the voluntary sector, managers have little choice but to become directly involved in evaluation. While this is also likely to be the case in art museums, Cousins et al. (2008) and Cousins, Goh, Elliott, Aubry et al. (2014) do not make clear which organizational types are grouped within the non-government/voluntary sector categories. Because it is impossible to know if art museums were sampled as part of these authors’ studies, gaining an understanding of evaluation capacity within Canadian art museums necessitates an independent empirical investigation as contextual considerations across the non-government sector are likely to be quite distinct. This notion is supported by research conducted by Carman and Fredericks (2010), who assert that nonprofit organizations are not a homogenous group.

Cousins, Bourgeois, and Associates (2014) applied the Conceptual Framework of the Capacity to Do and Use Evaluation (Cousins et al. 2008; Cousins, Goh, Elliott, & Bourgeois, 2014) to their multiple case study of EC in eight Canadian organizations representing a range of sectors (government: 4, voluntary: 3, education: 1). Analyses across the eight organizations comprising Cousins, Bourgeois et al.’s (2014) multiple case study revealed that most of the organizations in their sample were in a developmental mode with respect to their capacity to do evaluation (Cousins & Bourgeois, 2014). The two organizations that excelled in this capacity fell under the umbrella of the Canadian federal government and had both stand-alone evaluation units and annual budgets of $170 million and $100 billion respectively. A federal government organization and a para-governmental organization, both with evaluation units and annual
budgets of over $100 million, were among those organizations operating at higher levels of capacity to use evaluation.

The researchers (Cousins & Bourgeois, 2014) found EC to be most developed in organizations with leadership committed to evaluation, an organizational culture of learning, resident expertise in social science research methods, and well-established evaluation units. Such organizations also tended to forge evaluative partnerships (e.g., with consulting firms or universities), engage in direct ECB activities, involve staff, senior decision makers, and external stakeholders in evaluation processes, and use evaluation tools tailored to the needs of their organizations. While these findings add to the field’s understanding of EC and the factors that shape it, Cousins and Bourgeois (2014) noted that, because data collection and reporting for the individual case organizations took place in 2007, their study was already dated at the time of publication. A longitudinal research design, they acknowledge, could have been used to describe how these organizations’ EC might have changed or evolved over time.

**The Conceptual Model for the Measurement of EC (Nielsen, Lemire, & Skov, 2011)**

Nielsen, Lemire, and Skov (2011) conceptualize another EC framework derived from Harold Leavitt’s organizational model (Leavitt, 1978) and grounded in both the EC and ECB literature base. The authors’ framework is divided into two overall categories: *evaluation demand* (capacity to use evaluation) and *evaluation supply* (capacity to do evaluation). The demand for evaluation in an organization is conceived as driven by two sub-divided dimensions: *objectives*, and *structures and processes*. *Objectives* are understood as the purposes for which evaluations are conducted in public organizations. Both how these purposes are formalized (i.e., evaluation policies, guidelines, and annual plans) and how evaluation practices and findings are utilized within an organization are “indicative of an organization’s intentionality with respect to
Structures and processes refer to the application of evaluative knowledge in the functioning of an organization. This includes the financial resources allocated to the evaluation function and its organizational location (i.e., the executive, audit, or in-line branch of municipal government), the actual evaluation work of the people in the evaluation function, and the extent of an organization’s evaluation activity.

The second category, evaluation supply, consists of two sub-divided dimensions or building blocks for designing and implementing evaluations: human capital, and technology. Human capital pertains to the skills of those tasked with evaluation work, including their level of formal education and formal evaluation training as well as their practical evaluation experience and overall volume of other knowledge production. Broadly defined, technology applies to the means and resources that enable evaluation to be carried out, including evaluation models and data collection techniques as well as qualitative and quantitative data analysis software. Adequate technology was assumed to be in place in what the researchers term the “resource-rich Danish public sector” (p. 327).

Nielsen et al. (2011) empirically tested their framework in a nationwide survey of 287 evaluators working in Danish municipal governments (i.e., those who within the previous year had been responsible for implementing one or more evaluations in their respective municipalities). Confirmatory factor analysis (CFA) was conducted to determine whether or not the demand and supply dimensions of their framework were in fact expressed by their constituent dimensions. The CFA yielded results that supported the authors’ framework as a promising approach to measuring EC in Danish public sector organizations. However, their examination also revealed that evaluation supply was only moderately correlated with technology and human capital, indicating that further refinement of the supply side of their framework was needed. The
researchers’ study also indicated that Danish municipal governments were primarily found to use evaluations to attribute impacts and facilitate organizational learning with the highest-ranking municipalities characterized by a higher demand than supply score. Moreover, program evaluations were found to take place within a larger body of knowledge production that included results-based management, competence measurement, and cost-efficiency analysis. As such, the researchers remind us that “the boundaries between the craft of evaluation and other research-based forms of knowledge production are blurry” (p. 339). Therefore, in art museums the capacity to do and use program evaluation must be seen in relation to these institutions’ capacity to do and use a wide range of visitor studies.

Reflecting on their study, Nielsen et al. (2011) identified several limitations to their methodology. First, a small number of survey items (35–40) prevented the researchers from factoring out low-performing items in keeping with ideal instrument development procedures. Second, the number of respondents was not adequate to carry out a robust CFA and test the relationship between “the specific subcomponents and their relative contribution to the four sub-dimensions” (p. 332) in their model. Third, replicating the study with multiple samples would have helped to establish the reliability of the researchers’ results. By testing their model, Nielsen et al. (2011) were also able to identify several sub-dimensions that, having been omitted, served to weaken their framework. For example, a sub-dimension concerned with internal and external contextual pressures could have helped illuminate the conditions that impede or strengthen EC in Danish municipal governments. Sub-dimensions covering the role of evaluation culture and the role of knowledge management were also identified for inclusion in subsequent revisions of the authors’ framework.
The Synthesis Model of Evaluation Capacity (Taylor-Ritzler et al., 2013) adapted from the Suarez-Balcazar et al. (2010) Cultural and Contextual Framework of ECB

While the studies previously discussed are based on conceptual frameworks designed to describe and measure the key components of EC in the public sector, Suarez-Balcazar et al. (2010) presented a framework intended to build the capacity of service providers working in American disability-related community-based organizations (CBOs)—namely nonprofit organizations serving individuals with disabilities, centres for independent living as well as vocational rehabilitation offices. As a conceptual framework designed specifically for use in the nonprofit sector, Suarez-Balcazar et al.’s (2010) Cultural and Contextual Framework of ECB marks an important contribution to the evaluation field. Developed through a systematic review of the ECB literature and drawing on Bronfenbrenner’s (1967) ecological work, the authors’ framework outlines two central inter-acting factors assumed to influence ECB: individual factors and organizational factors. Individual factors are conceptualized as including personal readiness (i.e., willingness to partake in and motivation to learn about evaluation), competence (e.g., evaluative knowledge and skills), and personal cultural and contextual factors such as cultural competence and contextual awareness. Organizational factors consist of leadership that is supportive of evaluation, a learning climate that fosters evaluative thinking, and adequate resources to support evaluation in addition to both organizational context (e.g., history, environment, type, relationship with community, ways of operating) and organizational culture (e.g., how things are done, how members interact, traditions, norms, values).

In a 2013 study, Taylor-Ritzler, Suarez-Balcazar, García-Iriarte, Henry, and Balcazar report on the development and validation of the Evaluation Capacity Assessment Instrument (ECAI), a 68-item tool for measuring EC in nonprofit organizations based on both an abridged
and adapted version of the Suarez-Balcazar et al. (2010) framework heretofore referred to as The Synthesis Model of Evaluation Capacity. Specifically, the authors omitted all the contextual and cultural factors (i.e., organizational context, organizational culture, and personal cultural and contextual factors) present in Suarez-Balcazar et al.’s 2010, Cultural and Contextual Framework of ECB from their synthesis model while adding a new dimension: evaluation capacity outcomes which they further parsed out into the sub-dimensions of mainstreaming and use of evaluation findings.

After pilot testing the ECAI with four staff members from different nonprofits, Taylor-Ritzler et al. (2013) used the tool to survey 169 nonprofit social service agencies in the Chicago metropolitan area. Survey data were analyzed using a two-step analytic process that included both a confirmatory factor analysis (CFA) and structural equation modeling (SEM). CFA was conducted to test the validity of the ECAI by assessing the items included on each factor in the instrument while SEM was used to examine the relationships among the factors and outcomes delineated in the Taylor-Ritzler et al. (2013) synthesis model. Results from these analyses were found to provide empirical support for both the Synthesis Model of Evaluation Capacity and the use of the ECAI for measuring EC in nonprofit organizations. Additionally, the researchers’ findings showed that organizational factors affected the development of EC among individual staff, suggesting that ECB efforts should focus on organizational factors such as leadership, learning culture, and resources before targeting individual factors. This observation is consistent with those made by other researchers (e.g., Cousins et al., 2004; Preskill & Boyle, 2008).

Although Taylor-Ritzler et al.’s (2013) study provides statistical evidence of the importance of organizational factors in understanding EC, the researchers admitted that certain limitations related to the study sample, the methodology used, and the simplicity of the abridged
model tested should be considered when interpreting their findings. For instance, in addition to the fact that the relatively small study sample was restricted to one area of the United States, the study did not assess the role of respondents within the organizations sampled. As such, it is not clear from the researchers’ data whether and to what extent respondents were directly responsible for evaluation. Additionally, given the fact that the same sample was used for both the CFA and SEM analyses, further evidence from other participant groups and regions would be useful to better understand the validity of this framework across contexts. As identified by the researchers, future studies should sample multiple participants from different organizations over time to estimate both within- and between-organization variance in their capacity for evaluation, and ascertain the degree to which ECB efforts result in improvements in EC.

Taylor-Ritzler et al. (2013) likewise noted that the Synthesis Model of Evaluation Capacity does not include several important EC constructs identified in the evaluation literature. Process use (Patton, 2008), for one, is absent from the authors’ model. The researchers also acknowledge that the fact that the synthesis model does not explicitly consider organizational culture and context is problematic and recognize that instating contextual factors such as size, the role of evaluators within the organization, and both internal and external demand for evaluation in addition to cultural factors related to the organization, its programming, and the participants it serves would increase the validity of the model. A framework of EC in art museums should therefore reflect the complex contextual and cultural factors specific to this distinct sector (e.g., the relative importance and inclusion of education in the museum’s mission and the location of the education function within the hierarchy of the museum).
Comparison of Frameworks

The preceding studies provide an initial foundation of empirical evidence that supports the conceptualization and measurement of evaluation capacity (EC) in both government and nonprofit organizations. While there is consensus among scholars that EC is a complex multimensional construct, the conceptual frameworks published to date describe and measure a wide array of dimensions and sub-dimensions—some similar and some different. Below, I compare these frameworks’ constituent components to obtain a more nuanced understanding of how the dimensions of EC have been conceptualized in the literature to date and to identify those dimensions that may travel to the art museum context. Specifically, I examine (a) overlapping constructs and their location, (b) emphasis on and inclusion of factors and/or conditions influencing EC, (c) the inclusion and conceptualization of EC/ECB outcomes, (d) relationships among constructs, (e) intended use, and (f) empirical basis.

Overlap of constructs. Most of the conceptual frameworks in my review share at least some fundamental constructs. These include evaluative knowledge and skills, the capacity to use evaluation, ECB, resources and supports, evaluation activity, leadership, culture, and stakeholder involvement in evaluation. Notwithstanding this, the precise location of these constructs (i.e., where they are situated in each given framework) in addition to how they are framed and elaborated on differs based on the framework in question.

Knowledge and skills. All four of the EC frameworks published to date include evaluative knowledge and skills. The construct is captured under the evaluation logic and technical skills sub-dimension falling under the human resources dimension in Bourgeois and Cousins’ (2013) profile framework, the capacity to do evaluation dimension in Cousins et al.’s

**Capacity to use evaluation.** All four of the frameworks reviewed reference the capacity to use evaluation to varying extents. The use of evaluation *findings* and *process use* appear as separate stand-alone sub-dimensions under the *learning benefits* dimension in Bourgeois and Cousins’ (2013) profile framework and are subsumed under the *capacity to use* dimension in Cousins et al.’s (2008) conceptual framework. While both the use of *findings* and *practices* are included under the *objectives* dimension in Nielsen et al.’s (2011) conceptual model, only the capacity to use evaluation *findings* comprises the capacity to use evaluation in Taylor-Ritzler et al.’s (2013) synthesis model. Moreover, while Bourgeois and Cousins (2013), Cousins et al. (2008), and Nielsen et al. (2011) all unpack these constructs in their respective frameworks, Taylor-Ritzler et al. (2013) merely list *use of findings* as an outcome in their synthesis model.

**ECB.** *ECB* is included in all four of the frameworks reviewed. Labeled *professional development*, the construct appears as a stand-alone sub-dimension under the *human resources* dimension in Bourgeois and Cousins’ (2013) profile framework, as a stand-alone dimension (i.e., *sources of knowledge, skills and activities*) in Cousins’ et al.’s (2008) conceptual framework, a sub-dimension falling under the *human capital* dimension in Nielsen et al.’s (2011) conceptual model, and is subsumed under the *competence* dimension in Taylor-Ritzler et al.’s (2013) synthesis model.

**Resources and supports.** Resources and supports feature in all four of the frameworks reviewed. While these constructs appear to be parsed out among all the three dimensions comprising the *capacity to do* category in Bourgeois and Cousins’ (2013) profile framework (i.e., staffing, budget, and external supports) and the two categories comprising Nielsen et al.’s (2011)
conceptual model (i.e., funding for evaluation and data analysis software), the same constructs are captured under the *organizational support structures* dimension in Cousins et al.’s (2008) conceptual framework, and the *resources and supports* factor in Taylor-Ritzler et al.’s (2013) synthesis model.

**Evaluation activity.** Three of the four frameworks reviewed describe and measure an organization’s *evaluation activity* (e.g., nature, type, breadth, etc.). While in Bourgeois and Cousins’ (2013) profile framework, the construct once again appears to be parsed out among the three dimensions comprising the authors’ *capacity to do* category, it is captured under the *evaluative inquiry* dimension in Cousins et al.’s (2008) conceptual framework and the *structures and processes* dimension in Nielsen et al.’s (2011) conceptual model. In contrast, Taylor-Ritzler et al. (2013) do not explicitly refer to evaluation activity in their synthesis model. Additionally, two of the frameworks reviewed (Bourgeois & Cousins, 2013; Nielsen et al., 2011) take a broad view of *evaluation activity* by acknowledging that the production of evaluations occurs within a larger process of knowledge production in organizations.

**Leadership.** *Leadership* is included in three of the four frameworks reviewed. The construct appears as a stand-alone sub-dimension under the *human resources* dimension in Bourgeois and Cousins’ (2013) profile framework, an *organizational factor* in Taylor-Ritzler et al.’s (2013) synthesis model, and is included under the *organizational learning capacity* dimension in Cousins et al.’s (2008) conceptual framework. Nielsen et al. (2011) purposefully omit this construct from their model as the authors regard leadership as “being more an aspect of the building process than of actual EC” (p. 339).

**Culture.** Culture is included in three of the four frameworks reviewed. In Bourgeois and Cousins’ (2013) profile framework, organizational culture is referred to under the *organizational*
infrastructure sub-dimension falling within the organizational resources dimension. The construct is also referred to under the organizational learning capacity dimension in Cousins et al.’s (2008) conceptual framework and under the organizational factors category in Taylor-Ritzler et al.’s (2013) synthesis model. While neither organizational nor evaluation culture is covered in Nielsen et al.’s (2011) conceptual model, the authors plan to include the construct in future iterations of their model.

**Stakeholder involvement.** Two of the frameworks reviewed explicitly reference stakeholder involvement in evaluation. The construct is captured under the involvement in evaluation sub-dimension falling under the evaluation literacy dimension in Bourgeois and Cousins’ (2013) profile framework and the mediating conditions dimension in Cousins et al.’s (2008) conceptual framework. While the Suarez-Balcazar et al. (2010) Cultural and Contextual Framework of ECB, upon which the Taylor-Ritzler et al.’s (2013) synthesis model is based, includes stakeholder participation under its learning climate factor, the construct is not explicitly referred to in Taylor-Ritzler et al.’s (2013) synthesis model. Likewise, the construct does not appear to be included in either Nielsen et al.’s (2011) conceptual model.

**Factors and conditions influencing EC.** The four frameworks reviewed differ with respect to their emphasis on and inclusion of the factors and conditions that influence EC. While both Cousins et al.’s (2008) conceptual framework and Taylor-Ritzler et al.’s (2013) synthesis model focus explicitly on both constructs (i.e., by articulating these as stand-alone dimensions), factors and conditions seem to be much more embedded within the evaluation supply and demand categories in Nielsen et al.’s (2011) conceptual model and the capacity to do and use categories in Bourgeois and Cousins’ (2013) profile framework. Also, even though both Cousins et al. (2008) and Taylor-Ritzler et al. (2013) emphasize mediating factors and conditions, these
constructs are categorized differently in their respective frameworks—as organizational and individual factors in Taylor-Ritzler et al.’s (2013) synthesis model and as sources of knowledge skills and ability and organizational support structures in Cousins et al.’s (2008) conceptual framework.

The inclusion and conceptualization of EC/ECB outcomes. The two frameworks that explicitly include EC/ECB outcomes as stand-alone dimensions conceptualize these constructs differently. For example, Taylor-Ritzler et al. (2013) position both mainstreaming and use of findings as EC outcomes in their synthesis model while Cousins et al. (2008) position organizational learning capacity and organizational consequences as evaluation consequences in their conceptual framework. These conceptual differences likely stem from varying understandings of ECB’s ultimate purpose whether it be to generate high quality evaluation studies and make use of their findings (as in the former case) or to use evaluative knowledge to improve organizational effectiveness (as in the case of the latter framework). Bourgeois and Cousins’ (2013) profile framework and Nielsen et al.’s (2011) conceptual model do not explicitly include EC/ECB outcomes as stand-alone dimensions. However, given the organizing structure of Bourgeois and Cousins’ (2013) framework, it is implied that the capacity to do and use are indicators of an organization’s capacity for evaluation. Likewise, Nielsen et al.’s (2011) conceptual model implies that maximum EC occurs when the four dimensions in their framework are present in an organization.

Suggested links among constructs. The four frameworks reviewed elaborate on the relationships among their constituent components (i.e., categories, dimensions, sub-dimensions) to varying degrees. The Taylor-Ritzler et al. (2013) synthesis model, for instance, both identifies the outcomes of EC and specifies how individual and organizational factors relate to these
outcomes. Specifically, the model posits that organizational factors that are assumed to facilitate and/or hinder the transfer of individual learning into an organization’s practices while processes mediate the relationship between individual factors and sustainable evaluation practice. The Cousins et al. (2008) conceptual model considers the nature of EC and both identifies the consequences to which it might lead and the antecedent conditions that help to shape it. Nielsen et al. (2011), on the other hand, reason that an organization’s EC is concerned with both the supply and the demand side within an organization and is linked to its objectives, structure and processes, technology, and human capital, but do not explain the relationship between the supply and demand dimensions in their model. Likewise, although it is implied in Bourgeois and Cousins’s (2013) profile framework that human and organizational resources as well as evaluation planning and activities impact on an organization’s capacity to do evaluation, while evaluation literacy and integration with organizational decision making as well as learning benefits impact its capacity to use it, the authors do not specify how the dimensions and sub-dimensions in their framework are interrelated.

**Intended uses.** The purposes of the four frameworks reviewed and their intended uses differ somewhat. The Bourgeois and Cousins (2013) profile framework, for example, is positioned as a research tool for profiling EC in Canadian government organizations and helping practitioners identify capacity areas in need of improvement as well as a self-assessment tool for practitioners to use to reflect on and plan their ECB activities. While not conceptualized as profile tools, Nielsen et al.’s (2011) conceptual model and Taylor-Ritzler et al.’s (2013) synthesis model are also positioned as both (a) diagnostic tools for baseline or pretest assessment of EC (i.e., before ECB activities are initiated) aimed at informing what kind of capacity should be built and (b) tools for endline or posttest assessment of EC (i.e., after ECB activities are completed)
aimed at measuring changes in capacity to determine the success of capacity development efforts. In addition, the Taylor-Ritzler et al. (2013) synthesis model is also positioned as a monitoring tool to be used over time “to assess the sustainability of evaluation capacity within organizations and inform ongoing ECB efforts” (p. 202). In contrast to these frameworks, the Cousins et al. (2008) conceptual framework is the only framework reviewed that is intended to be used solely by evaluation researchers to conduct posttest assessments of EC and contribute to the literature on ECB.

**Empirical basis.** Finally, the empirical bases also differ across frameworks. Some of the frameworks reviewed are based on qualitative research designs (e.g., Bourgeois & Cousins, 2013) while others are based on quantitative ones (i.e., Nielsen et al. 2010; Ritzler et al., 2013) making it possible to infer statistical generalizability (Yin, 2009) from them. In all four cases, the frameworks focus on a specific sector of activity. Some of the frameworks have gained empirical support within the context of Canadian (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008) and Danish (i.e., Nielsen et al., 2011) government sectors. Another has gained empirical support within the context of the American nonprofit sector (i.e., Taylor-Ritzler et al., 2013). The empirical bases for these frameworks make them well suited to describing and measuring EC within these specific organizational environments, but less suited to others. Their meaningful use in other organizations would require more extensive development and validation.

**Summary and Limitations of the Frameworks Reviewed**

The four frameworks reviewed (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen et al., 2011; Taylor-Ritzler et al., 2013) may share some fundamental dimensions, however, their comparison also illustrates the fact that, akin to the absence of an agreed-upon definition of evaluation capacity (EC), consensus is lacking in the field about the precise
dimensions and sub-dimensions that comprise organizations’ capacity for evaluation. Rather, expositions of EC, as Nielsen et al. (2011) point out, are characterized by “widespread conceptual pluralism” (p. 324). While the idea of establishing both a general definition and common framework of EC in the name of conceptual clarity intrigues some scholars in the evaluation community (e.g., Suarez-Balcazar et al., 2010; Taylor-Ritzler et al., 2013), Taut (2007) asserts that a range of conceptualizations is necessary to accommodate contextual differences in evaluation. Bourgeois, Toews, Whynot, and Lamarche (2013) second this notion by asserting that the availability and diversity of such frameworks point to the need for adaptation rather than generalization.

As implied, both above and elsewhere in this paper, the main limitation of the conceptual frameworks published to date is the context within which they were developed and have gained empirical support. More specifically, while these frameworks add to the field’s understanding of evaluation capacity in government settings and disability-related community-based organizations, they are limited in their abilities to support research on EC in art museums as many of their constituent dimensions (i.e., results-management orientation, reward systems, peer reviews and meta-evaluations, systematic reviews of evaluation units, etc.) do not translate beyond the organizational type and sector for which they were originally designed. Bourgeois, Whynot, and Thériault (2015) explained, “Part of the difficulty in using such models in different organizations lies in the applicability of certain dimensions from one organizational type or sector to another” (p. 48).

A prime example of such a dimension is the *results-based management orientation* dimension in Bourgeois and Cousins’ (2013) profile framework. This orientation is reflective of federally-driven evaluation policies that align the evaluation function in Canadian government
with new public management\textsuperscript{10} reforms focused on managing for results (Treasury Board of Canada Secretariat, 2000, 2006). Since the overwhelming majority of Canadian art museums are not governed by such policies or reforms, they are unlikely to embrace these evaluation systems. In addition to this, several of the framework’s other dimensions (e.g., human resources, organizational resources, evaluation planning and activities, etc.) assume the presence of evaluation units and speak almost exclusively about the work of evaluators employed within such units. While stand-alone evaluation departments may be ubiquitous in Canadian government departments and agencies, they are largely non-existent in Canadian art museums. Larger art museums may have individual staff dedicated to the evaluation function, but in most art museums the task of evaluating educational programs falls to program staff who, unlike non-evaluator government employees, must conduct evaluations themselves due to resource constraints.

A number of other dimensions and sub-dimensions present in Cousins et al.’s (2008) conceptual framework pose further limitations. The capacity to do evaluation, for instance, while highly relevant to art museums, includes formal reward systems for the use of performance information which, in my experience, is an unheard-of practice in art museums. Rather than reflecting the types of evaluation activities that occur in art museums (e.g., museum exit surveys, sweep rates, visitor tracking studies, exhibition evaluations, etc.), evaluative inquiry in this framework pertains to government-centric evaluation activities such as audits and environmental scans. Mediating conditions likewise focuses too narrowly on factors (i.e., involvement of users and evaluator credibility) that influence the use of evaluation in the context of the Canadian federal government. In this context, evaluations are frequently informed by user input on account

\textsuperscript{10} The term new public management refers to “a renewed stress on the importance of management” in public administration (Hood, 2001, p. 12553).
of the arms-length relationship that exists between evaluators and various stakeholder groups. In art museums, the evaluator and the user are, more often than not, one and the same. Given that these individuals are unlikely to be trained or credentialed evaluators, they may not be familiar with the professional standards of practice (e.g., Competencies for Canadian Evaluation Practice [Canadian Evaluation Society, 2019]; the Program Evaluation Standards [Yarbrough, Shulha, Hopson, & Caruthers, 2011]) to which evaluators in government settings are required to subscribe. Where present, art museum evaluators are more likely to use the Visitor Studies Professional Competencies (VSA, 2019b) to guide their work. Regardless of who is doing the evaluating or what standards of practice they adhere to, program evaluations in art museums are also unlikely to be supported by peer reviews or meta-evaluations, as may be the case in government.

As can be seen from some of the limitations reviewed above, these models’ empirical groundings pose certain limitations in terms of their applicability to the study of EC in art museums. Clearly, a conceptual framework (i.e., one that draws on and adapts relevant dimensions from others) is still needed to guide research on EC in the art museum sector. In the section that follows, I present such a framework.

A Conceptual Framework of Evaluation Capacity in Art Museums

Based on an analysis of existing evaluation capacity (EC) frameworks (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011; Suarez-Balcazar et al., 2010; Taylor-Ritzler et al., 2013) and informed by the museum literature as well as my own exploratory research into program evaluation in the sector (Chalas, 2016) and by my prior professional museum experiences, I have devised an initial Conceptual Framework of Evaluation Capacity in Art Museums that articulates the key dimensions of EC as they could be observed in
art museums. Following the above frameworks, a multi-dimensional framework has been conceptualized for this sector. The framework consists of two main categories and seven subdivided dimensions as outlined in Figure 3 below.

Figure 3. The conceptual framework of evaluation capacity in art museums

In brief, the first category describes four factors that are posited to influence evaluation practice and capacity in art museums in addition to the extent to which art museums are likely to engage in direct ECB activity. This category is comprised of the following four factors or dimensions: context, culture, demand for evaluation, and resources and supports. Context refers to a museum’s size, location, age, visitor demographics, sources of funding, mandate, and structure. Culture encompasses both the museum’s institutional and evaluation culture. Demand for evaluation focuses on both internal and external expectations for evaluation information. Finally, resources and supports include time as well as human, financial, technological, and
other resources.

The second category in my framework consists of three dimensions conceptualized as interrelated capacity indicators—*capacity to do and use, evaluation practice, and direct ECB activity*. Grouped under the *capacity to do and use* are two sub-dimensions: *the capacity to conduct visitor studies and use visitor studies information* and *the capacity to do and use program evaluations*. Because understanding an art museum’s capacity to do and use program evaluations requires an understanding of a museum’s overall capacity for evaluation, these sub-dimensions are envisioned as nested constructs, with the latter being situated within the former. Likewise, *evaluation practice* is also divided into two nested sub-dimensions with *program evaluation activity* being situated within a museum’s overall *visitor studies activity* for the same reason as articulated above.

My previous research (Chalas, 2016) has revealed that discrepancies in both an art museum’s capacity to do and use evaluation and its evaluation practices at the organizational vs. departmental levels can be indicative of the systemic marginalization of educational endeavours within such institutions. In other words, many art museums still choose to allocate their resources to areas of visitor studies inquiry such as exhibition evaluation over the evaluation of educational programs, despite museum education’s rise in stature over the last few decades. Comparing an art museum’s capacity to do and use program evaluations with its capacity to do and use a wider range of visitor studies, as well as comparing its program evaluation activity to its overall visitor studies activity is therefore crucial to both ascertaining an accurate picture of a museum’s capacity for evaluation and understanding its evaluation culture. This notion is consistent with those brought forward by others (e.g., Nielsen et al., 2011) who assert that evaluations are one of several forms of knowledge produced to inform decision making in organizations.
The last dimension in my framework focuses on determining the presence of *direct ECB activities*. Moreover, akin to Taylor-Ritzler et al.’s (2013) and Cousins et al.’s (2008) frameworks, the *Conceptual Framework of Evaluation Capacity in Art Museums* makes explicit the hypothesized relationship among the three dimensions conceived of as capacity indicators—*capacity to do and use, evaluation practice, and direct ECB activity*. It is expected, for instance, that a higher level of EC would be associated with a museum’s use of effective ECB processes. A more developed capacity for evaluation, in turn, is assumed to lead to higher quality evaluation studies and the mainstreaming of evaluation practices into an art museum’s operations. Engaging in evaluation activities is presumed to feed back into a museum’s capacity to both conduct research and evaluation and make effective use of study findings. A more detailed description of each of the framework’s seven dimensions follows and is situated in other published frameworks.

**Context.** It is well understood that evaluation takes place within a particular context (Compton, Baizerman, & Stockdill, 2002). Several evaluation researchers (e.g., Cousins et al., 2004; Preskill & Boyle, 2008) have noted that context can have a significant impact on an organization’s capacity for evaluation. Like Suarez-Balcazar et al.’s (2010) cultural and contextual framework, my framework also pays particular attention to context.

Specifically, this dimension focuses on such contextual characteristics as a museum’s *size* (e.g., annual operating budget), *type* (e.g., nonprofit vs. Canadian crown corporation), *age* or level of maturity, *geographical location, visitor demographics*, and *sources of funding* in addition to its broad *mandate* as expressed by the range of activities it supports, from the artworks it acquires and exhibits to the educational programs it offers its diverse publics. The dimension also includes a museum’s institutional *structure*—especially as it pertains to the
location and composition of the education function within, what Samis and Michaelson (2017) refer to as, the hierarchy of the museum. These are all important contextual factors that are likely to impact on an art museum’s capacity for evaluation.

Ascertaining these museum characteristics can be done in part through a document analysis (e.g., annual reports). General questions that enable investigation into context are: What is the full extent of the programs offered by educational departments and their associated target audiences? What is the museum’s organizational and funding structure? What policies (if applicable) govern program evaluation in art museums?

**Culture.** Museums, like other organizations, are also characterized by their idiosyncratic cultures (i.e., the customs, traditions, and competing values and assumptions that shape behavior in museums) (Davies, Paton, & O’Sullivan, 2013). This dimension thus focuses on diagnosing both an art museum’s *institutional* culture and its *evaluation* culture.

Aspects of a museum’s *institutional culture* pertinent to my study include the value placed on museum education as expressed by the working relationship between education and curatorial staff and the degree of the presence of a hierarchical divide between curatorial and educational endeavours within a given museum. *Evaluation culture* refers to the extent that an art museum has embraced evaluative thinking and mainstreamed the evaluation process into its day-to-day operations. Examples of an evaluative mindset or *habit of mind* (Katz, Sutherland, & Earl, 2002) include leadership that demonstrates a commitment to evaluation in addition to both an awareness of the need for and the value of evaluation and an interest in evidence-informed visitor-centred decision making at the individual, group (i.e., departmental) and institutional levels. An art museum’s evaluation culture may also expressed by the presence or absence of an institution-wide evaluation policy as well as a dedicated visitor studies/evaluation function.
In conceptualizing this dimension, I draw primarily on the work of Suarez-Balcazar et al. (2010) who place primacy on culture in their cultural and contextual framework in addition to the work of Bourgeois and Cousins (2013), Cousins et al. (2008), and Taylor-Ritzler et al. (2013) who also include the construct in their respective frameworks. General questions that enable investigation into culture are: How do program staff describe the status and role of the education department within the museum? How do they characterize staff attitudes towards evaluation?

**Demand for evaluation.** Both visitor studies and program evaluation in art museums necessitate an effective demand for these types of activities. Bourgeois and Cousins (2013) spoke to this notion when they said, “External accountability requirements often create a demand for evaluation results” (p. 301). Much the same can be said of internal pressures for evaluation. By contrast with the other frameworks reviewed in this chapter (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011; Taylor-Ritzler et al., 2013; Suarez-Balcazar et al., 2010), demand for evaluation is conceptualized as a stand-alone dimension in my framework. This dimension aims to ascertain both the drivers of visitor studies and program evaluation in Canadian art museums (e.g., internal impetus vs. external pressure) and to better understand art museum funders’ accountability expectations in the Canadian context. The dimension also probes the challenges posed by funder-mandated accountability demands.

General questions that enable investigation into this dimension include: What was the main reason the museum decided to engage in evaluation (e.g., at the request of the board of directors, or to fulfill funder requirements)? To what degree do the agencies and foundations that fund educational programs at the museum make evaluation mandatory? What kind of information do funders expect (e.g., evaluations that report on the outcomes of the programs vs. those that report on outputs)?
**Resources and supports.** Support in the form of *time, human, financial,* and *technological* resources is crucial to sustained visitor study including educational program evaluation in art museums. Drawing on similar dimensions present in the frameworks reviewed in this chapter (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011; Taylor-Ritzler et al., 2013; Suarez-Balcazar et al., 2010), this dimension addresses not only the location of the evaluation and visitor studies functions within the museum, but also the amount of time and other resources allocated to these functions. Specifically, the *human resources* sub-dimension examines the organizational role of those charged with research and evaluation at the museum (e.g., art museum educator with evaluation responsibilities vs. internal evaluator vs. external consultant). Related to this, *time* pertains to the percentage of time that museum staff spend engaged in conducting or overseeing visitor studies and/or evaluation activities compared to the amount of time they dedicate to their other job duties (if applicable). *Financial resources* probe both the internal and external sources of funds allocated to visitor studies and evaluation, with particular emphasis placed on determining the degree to which Canadian funders’ demands for evaluation are accompanied by concomitant financial resources. This sub-dimension also encompasses the financial support (if any) that funders make available for ECB activities.

Other funder *supports* captured by this dimension include the provision of advice on evaluation (e.g., appropriate evaluation measures, the development of funding formulas for evaluation activities), technical assistance, training support aimed at building museum practitioners’ evaluative knowledge and skills in addition to evaluation tools and resources. *Technology* refers to the availability of software used to compile, analyze, and manage
evaluation information. Lastly, this dimension also assesses an art museum’s perceived evaluation resource challenges and needs.

General questions that enable investigation into this dimension include: To what extent do educational program funders provide funding to support evaluation activities or allow project funds to be used for evaluation purposes? Who is responsible for conducting educational program evaluations at the museum, and what percentage of that person’s time is dedicated to the evaluation function?

**Capacity to do and use evaluation.** Akin to Cousins et al.’s (2008) conceptual framework, both the capacity to do and use evaluation are included in my framework albeit as one stand-alone dimension of the same name. In my conceptualization, this dimension refers to an art museums’ capacity to do and use program evaluations as compared to its overall capacity to conduct a wide range of visitor studies and make effective use of the information collected. More specifically, the capacity to do evaluation refers to the knowledge and skills required to conduct or oversee program evaluation activities in addition to the knowledge and skills required to carry out a wide range of visitor studies activities in art museums (e.g., timing and tracking studies, exhibition and museum exit interviews, visitor satisfaction surveys). Competence in program evaluation and visitor studies includes the ability to use knowledge and skills related to these areas to frame research and evaluation questions, plan studies, address privacy and ethical issues, design instruments, and select appropriate methods and approaches. It also involves knowledge and skills pertaining to data collection, analysis, interpretation, and reporting.

As in Cousins et al.’s (2008) conceptual framework, the capacity to use evaluation is reflected in both an art museum’s *use of findings* and *process use*. The *use of findings* encompasses instrumental uses (e.g., to inform exhibition or educational program decision
making), conceptual uses (e.g., to learn about a program’s strengths and weaknesses), and symbolic or persuasive uses (e.g., to increase the status of educational departments by promoting their work within the hierarchy of the museum and helping to attract funding support for new and expanded programming) (Patton, 2008). Coined by Patton (1997), process use refers to the extent that members of an organization benefit from engaging in the evaluation process. Process benefits can include the development of evaluative knowledge and skills and learning how to think evaluatively. This dimension also addresses perceived challenges to an art museum’s capacity for evaluation in addition to its professional development needs.

General questions that enable investigation into this dimension include: What visitor studies/evaluation-related training do those charged with evaluative responsibilities have? What is the greatest strength they bring to the evaluation of the museum’s education programs? What did they learn from carrying out their last evaluation/visitor study?

Evaluation practice. This dimension draws on three of the frameworks reviewed in this chapter that also describe and measure this construct (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen et al., 2011). In conceptualizing this dimension, I, like Bourgeois and Cousins (2013) and Nielsen et al. (2011) before me, also take a broad view of evaluation practice—by acknowledging that the production of evaluations occurs within a larger process of knowledge production in organizations. Evaluation practice is thus concerned with capturing the nature and extent of both the program evaluation and visitor studies activities occurring in art museums. More specifically, this dimension aims to ascertain what gets evaluated in art museums (e.g., school program vs. exhibition display) and which evaluation methods (e.g., informal staff meetings vs. formal evaluations, qualitative vs. quantitative) and approaches (e.g., outcome-based vs. participatory) get used by art museum sector practitioners. Implied in the
latter is both the degree to which museum practitioners informally learn about evaluation by doing it (i.e., indirect ECB) and the degree to which various stakeholders are involved in the evaluation process—especially in the case of educational and exhibition programs to which visitors may have contributed as co-creators.

Evaluation practice also looks at the type of evaluation information that gets collected (e.g., outputs vs. outcomes, educator-prescribed outcomes vs. visitor-defined outcomes) and how this information is analyzed, interpreted, and reported. Additional considerations include the number, frequency, and quality of the evaluations performed. Likewise included under this dimension are the challenges practitioners face when engaging in visitor studies and program evaluation activities in art museums and their needs for assistance. General questions that enable investigation into this dimension include: What specific programs or projects did practitioners evaluate? How did they measure program impact?

ECB. Like all four of the frameworks reviewed in this chapter (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011; Taylor-Ritzler et al., 2013), my framework also includes direct ECB activity as an important construct. This last dimension seeks both to capture whether or not art museums are actively engaged in building their evaluative capacities and to determine the specific activities used to do so. This includes intentional capacity building initiatives directed at enhancing evaluative knowledge and skills (e.g., courses and workshops) in addition to other approaches to capacity development in museums such as forging museum-university partnerships (Owen & Visscher, 2015) and forming evaluation networks for museum professionals (Adams, 2012; Steele-Inama, 2015). Also captured by this dimension are art museum practitioners’ perceived ECB needs and challenges. General questions that enable investigation into this dimension include: What strategies has the
museum used to enhance its internal evaluation efforts or build the evaluation capacity of its staff? To what extent were these efforts successful?

Differentiating the Proposed Framework

While the Conceptual Framework of Evaluation Capacity in Art Museums features some of the same dimensions (i.e., capacity to do and use, evaluation practice, resources and supports, etc.) as identified in the EC frameworks previously reviewed in this chapter, it is distinguished from these frameworks in five important ways. Specifically, the framework

- pays particular attention to culture and context in arts museums,
- includes constructs that are purposefully embedded in the art museum sector,
- focuses on an art museum’s capacity to evaluate its educational programs as compared to its overall capacity to conduct other forms of visitor studies inquiry and explicitly situates these two capacities as nested constructs,
- assumes a mix of practitioner and evaluator involvement in evaluation unlike most of the other frameworks reviewed (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011), and
- addresses both the challenges art museums encounter related to each framework dimension and the resources they need to increase their capacity for evaluation.

Given the above, the framework is uniquely suited to describing evaluation practice and diagnosing EC in the art museum sector. In this way, the Conceptual Framework of Evaluation Capacity in Art Museums addresses the first purpose of this study and makes an original contribution to the theoretical literature base related to the conceptualization of EC.
Uses of the Proposed Framework

Following Taylor-Ritzler et al.’s (2013) framework, my framework is envisioned as a diagnostic, post-ECB intervention, and monitoring tool for use by evaluation researchers and museum practitioners alike. As a tool for baseline diagnosis of an art museum’s capacity for evaluation (i.e., pre-ECB intervention), it could be useful in identifying ECB activities aimed at improving low capacity areas. While not intended to assess the capacity building process itself or the practices employed within it, the framework could also be used throughout the ECB process for both monitoring purposes (i.e., midpoint assessment aimed at informing ongoing ECB efforts) and summative assessment of EC (i.e., post-intervention assessment aimed at ascertaining the degree to which ECB efforts have led to improvements in capacity). The measurement of EC is therefore envisioned as an ongoing activity within the ECB process as depicted in Figure 4 below.

![Diagram of ECB process](image-url)

Figure 4. The location of EC diagnosis within the ECB process

In the present study, the Conceptual Framework of Evaluation Capacity in Art Museums served to bound my research into EC in the Canadian art museum sector. Specifically, the framework was used to guide instrument development as well as data collection, analysis, interpretation, and reporting in each of the study’s two phases. In the chapter that follows, I detail the ways in which I applied the framework in my research.
Chapter 3
Methodology

This study used a two-phase multiple-method qualitative research design to address the purposes of this research and answer the research questions. Phase One involved conducting in-depth interviews with key informants in a cross-section of art museums across Canada. Phase Two comprised case studies focused on two purposefully-selected art museums. Each of these two interrelated phases of data collection, analysis, and interpretation had a distinct focus described in greater detail below. This description is followed by a discussion of the strategies used to validate findings from both phases of the research. Next, an explanation of the ethical assurances that were put in place is provided.

Phase One: Interview Study

The first phase of my research involved conducting an interview study to establish an initial knowledge base on evaluation (i.e., practices, capacities, building efforts, influencing factors, needs, and challenges) in the Canadian art museum sector. Additionally, this phase of my study also sought to test the degree to which the Conceptual Framework of Evaluation Capacity in Art Museums could be considered an accurate or complete description of evaluation capacity (EC) in this context. The main research questions that framed this phase of the study were the following:

1. What are the key dimensions that describe EC in the Canadian art museum sector?
2. How is EC—specifically program evaluation capacity—manifested across Canada’s largest publicly owned art museums, and to what extent does this capacity vary from museum to museum?
Overall, the results of the qualitative interview study illuminated the current landscape by describing Canadian art museums’ capacity to evaluate their educational programs and were used to both refine my conceptual framework and identify two art museums operating at the most developed level of capacity for in-depth case study.

Selection and recruitment of phase one participants. Phase One employed a key informant purposeful sampling strategy (Patton, 2015) to identify interview participants working in publicly owned art museums across Canada, namely art museum practitioners with knowledge of and/or primary responsibility for either overseeing or conducting evaluations of educational programs in their respective institutions. Patton (2015) pointed out that key informants are “especially important sources on specialized issues” (p. 284). This phase of my study relied heavily on key informants because the topic of evaluation practice and capacity within the art museum sector is a highly specialized one. This sampling strategy thus enabled me to select informants capable of providing valuable insights into the key dimensions of EC in this context and allowed me to reveal important patterns, describe between-museum variations in capacity, and develop recommendations for improving the program evaluation function across the sector.

Key informants were recruited from Canada’s largest publicly owned art museums of which there are a total of fourteen across the nation as listed in Table 2 below. These fourteen art museums were identified through various museum networks (e.g., the Ontario Museum Association, the Association of Manitoba Museums, the British Columbia Museums Association) and met the criteria for inclusion. Specifically, they were all publicly owned art museums that

- had annual operating budgets of, at minimum, 2.5 million dollars;
- focused their missions on collecting, exhibiting, preserving, and educating about works of
visual art;

- supported stand-alone education departments that
  - offered an expansive schedule of both educational (e.g., school programs) and public/community programs (e.g., artist talks) for a diverse range of audiences, and
  - employed either a Head of Education who was supported by additional staff members or at least two education staff members working in coordinating or management roles; and

- received funding for their educational programs in the form of grants.

Given the above criteria, small and mid-sized regional art museums (e.g., the Art Gallery of Sudbury, the Dunlop Art Gallery, the Art Gallery of South Western Manitoba), university art museums (e.g., the McMaster Museum of Art, the University of Calgary Nickle Galleries, the Museum of Anthropology at the University of British Columbia), and museums with a secondary mandate related to the heritage sector (e.g., Glenbow, the Orillia Museum of Art and History, the Confederation Centre Art Gallery and Museum) were excluded from this interview study.

Previous research conducted in the nonprofit sector (e.g., Carman, 2009; Carman & Fredericks, 2010) suggests that larger nonprofit organizations and those that receive funding from the federal government have greater success when it comes to program evaluation and that interview respondents are more likely to be from such organizations. Because Canadian art museums are also registered nonprofit organizations, the decision to interview key informants working in Canada’s fourteen largest publicly owned art museums was made to increase the probability of participation as well as the probability that I would be gathering data from museums that actually conduct program evaluations (even if their capacity to do so varied). It
should be noted that while the table below shows considerable geographic spread (i.e., eight of ten Canadian provinces), it excludes the three territories (i.e., Northwest Territories, Nunavut, and Yukon) as art museums matching the above criteria do not exist in those locations.

Table 2

**Canadian art museums sampled**

<table>
<thead>
<tr>
<th>Museum</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art Gallery of Alberta</td>
<td>Edmonton, Alberta</td>
</tr>
<tr>
<td>Art Gallery of Greater Victoria</td>
<td>Victoria, British Columbia</td>
</tr>
<tr>
<td>Art Gallery of Ontario</td>
<td>Toronto, Ontario</td>
</tr>
<tr>
<td>Art Gallery of Nova Scotia</td>
<td>Halifax, Nova Scotia</td>
</tr>
<tr>
<td>Beaverbrook Art Gallery</td>
<td>Fredericton, New Brunswick</td>
</tr>
<tr>
<td>Hamilton Art Gallery</td>
<td>Hamilton, Ontario</td>
</tr>
<tr>
<td>Mackenzie Art Gallery</td>
<td>Regina, Saskatchewan</td>
</tr>
<tr>
<td>Montreal Museum of Fine Arts</td>
<td>Montreal, Quebec</td>
</tr>
<tr>
<td>Musée d'art contemporain de Montréal</td>
<td>Montreal, Quebec</td>
</tr>
<tr>
<td>Musée national des beaux-arts du Québec</td>
<td>Quebec City, Quebec</td>
</tr>
<tr>
<td>National Gallery of Canada</td>
<td>Ottawa Ontario</td>
</tr>
<tr>
<td>Remain Modern</td>
<td>Saskatoon, Saskatchewan</td>
</tr>
<tr>
<td>Vancouver Art Gallery</td>
<td>Vancouver, British Columbia</td>
</tr>
<tr>
<td>Winnipeg Art Gallery</td>
<td>Winnipeg, Manitoba</td>
</tr>
</tbody>
</table>

Recruitment of key informants took place over a three-month period beginning in August of 2017. Because art museum practitioners representing the program evaluation function in art museums are likely to be education staff, combined letters of invitation and consent forms (Appendix D) were initially sent, via email (see Appendix B for an example email script), to the Head of Education at each of the 14 museums selected for participation in the study. Follow-up phone calls (see Appendix C an example telephone script) were then made two weeks later. Where the Head of Education did not hold knowledge of or direct responsibility for either overseeing or conducting the evaluation of educational programs, s/he was asked to assist me with identifying a more suitable key informant within their institution (e.g., an internal
evaluator). A total of 13 art museum practitioners representing 11 art museums from six provinces across Canada agreed to participate in this phase of the study. Practitioners from the remaining three art museums did not respond to either initial or follow-up invitations for participation.

**Interviews.** In this study, I used a semi-structured qualitative interview approach (Patton, 2015) to develop an interview protocol consisting of a mix of open- and closed-ended questions. Appendix E provides the interview protocol used in the first phase. I chose this interview approach for three reasons. First, a multi-site interview scenario called for minimizing variation in the questions posed to interviewees to enable comparability of responses across participating art museums. Second, because museum practitioners are busy individuals, a highly-focused interview ensured that interviewee time was used efficiently. Lastly, the approach was selected because it served to facilitate both the organization and analysis of data. Patton (2015) explained, “If a standardized open-ended interview has been used, it is fairly easy to do cross-case or cross-interview analysis for each question in the interview” (p. 534).

When developing my interview protocol, I paid careful attention to the wording and sequencing of each question and wrote any prefatory statements, transitions, or standard probes to be used directly into the instrument itself. The development of my interview protocol was further guided by the dimensions present in the *Conceptual Framework of Evaluation Capacity in Art Museums*. As such, interview questions asked key informants to report on both their art museum’s contextual characteristics and extant evaluation culture. Practitioners were likewise asked about (a) what drove program evaluations at their museums (i.e., internal demands vs. external pressure), (b) the type and amount of evaluation resources and training supports that were available to them, (c) their current capacity to collect, reflect on, and incorporate various
forms of data into their educational program development processes, and (d) the specific evaluation activities they engaged in to document programmatic impacts. Additional questions probed key informants’ perceived challenges with evaluation and their needs for assistance. It is important to note that I did not ask key informants to directly report on their organizational contexts at this time, reserving the in-depth examination of this influencing factor for the second phase of the study.

As shown in Table 3, a total of 12 in-depth interviews (one being a group interview) of around 60 minutes in length were carried out with 13 key informants (i.e., 7 Heads of Education, 5 education staff members with responsibility for evaluation, and 1 internal researcher) representing the eleven art museums that agreed to participate in this phase of my study—a response rate of 79%. The overall purpose of the interview was reiterated to participants at its beginning, and, while a standardized interview format was used the majority of the time, I reserved the latter part of the interview to pursue select topics in greater depth or even to pose additional questions that were not originally anticipated in my instrument development. In this way, I was relying on a combined deductive-inductive approach to data collection.

All 12 interviews were conducted via telephone and recorded using a digital recorder. Patton (2015) pointed out that “doing all or some of your own interview transcriptions (instead of having them done by a transcriber) provides an opportunity to get immersed in the data, an experience that usually generates important insights” (p. 525). For this reason, I undertook transcription myself. The transcripts were subsequently checked for mistakes by interviewees (Gibbs, 2008) and edited to unite sections of a narrative that were either interrupted or veered off topic (Holliday, 2002).
Table 3

Overview of participants and data sources per museum

<table>
<thead>
<tr>
<th>Museum</th>
<th>Participants (13)</th>
<th>Data Sources (12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Education staff member with responsibility for evaluation</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>2</td>
<td>Head of Education + internal researcher</td>
<td>Two separate 60-minute interviews</td>
</tr>
<tr>
<td>3</td>
<td>Head of Education</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>4</td>
<td>Education staff member with responsibility for evaluation</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>5</td>
<td>Head of Education</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>6</td>
<td>Education staff member with responsibility for evaluation</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>7</td>
<td>Head of Education</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>8</td>
<td>Head of education + Education staff member with responsibility for evaluation</td>
<td>One 60-minute group interview</td>
</tr>
<tr>
<td>9</td>
<td>Head of education</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>10</td>
<td>Education staff member with responsibility for evaluation</td>
<td>One 60-minute interview</td>
</tr>
<tr>
<td>11</td>
<td>Head of education</td>
<td>One 60-minute interview</td>
</tr>
</tbody>
</table>

**Data analysis and reporting procedures.** I employed a framework approach to qualitative content analysis (Ritchie & Spencer, 1994, 2002) to manage, analyze, and interpret the interview data collected during Phase One. I selected this approach because it (a) helps in managing large qualitative data sets where obtaining a comprehensive review of all the data collected is desirable, (b) enables comparisons between and within interviews to be made, (c) allows for easy retrieval of text, (d) facilitates methodical treatment of all similar units of analysis, and (e) enhances the credibility of findings by making the analytic process transparent to others (Ritchie & Lewis, 2003; Ritchie, Lewis, McNaughton Nicholls, & Ormston, 2014; Ritchie & Spencer, 1994, 2002). The framework approach consists of five interconnected stages of analysis as described by Ritchie and Spencer (1994, 2002): familiarization, developing a coding system, coding, charting, and mapping and interpretation. Below, I describe how I moved through each stage of analysis.
Familiarization. After preparing the interview data for analysis in the manner described in the previous section, I familiarized myself with my data set by both re-listening to all the interviews and reading each individual interview transcript carefully to “obtain a general sense of the information and to reflect on its overall meaning” (Creswell, 2009, p. 185). While doing so, I simultaneously recorded my first impressions of the data and any thoughts I had about analysis in a research journal.

Developing a coding system. Because I based my data collection, analysis, interpretation, and reporting on the Conceptual Framework of Evaluation Capacity in Art Museums, I developed a “start list” of codes (Miles, Huberman, & Saldana, 2014, p. 81) that was predetermined by my conceptual framework. Specifically, this initial list (Appendix F) consisted of seven framework-driven master codes (e.g., Dimension: Evaluation Practice) and a further 36 sub-codes (e.g., Program Evaluation Activity: Stakeholder Involvement) and their associated abbreviations (e.g., ACT-SI) that were identical to both the dimensions and sub-dimensions in my framework and the various areas of questioning specified in my interview protocol (Appendix E). Together, the complete set of master and sub-codes on my start list constituted my working analytical framework.

Coding. Having developed my start list of codes, I systematically applied them to the interview data using a combined deductive-inductive approach. During deductive coding (Miles, Huberman, & Saldana, 2014), I read the transcripts line-by-line and manually assigned the abbreviated sub-codes from my start list (Appendix F) to corresponding segments of interview text. Following this deductive coding process, I manually open-coded (Strauss & Corbin, 1998) each interview transcript in an effort to uncover additional codes in the data that were not anticipated at the beginning of the study. The data-driven codes that emerged from this inductive
process were then used to refine my start list of *a priori* codes (see Appendix J for a final list of codes applied during Phase One). Therefore, while I used the *Conceptual Framework of Evaluation Capacity in Art Museums* to guide analysis, its use did not preclude the emergence of inductive themes.

**Charting.** After coding individual transcripts and refining my code list, I synthesized my entire data set into a structured matrix coding chart (see Appendix G for examples) to “build up a picture of the data as a whole” (Ritchie & Spencer, 2002, p. 317) in preparation for interpretation. I began this process by drawing up a chart for each of the master codes (i.e., dimensions) in my final code list. The charts were comprised of both rows (i.e., one for each sub code) and columns (i.e., one for quotations and transcript line references and another for text summaries). Next, I summarized each coded segment of interview text and then entered the resulting summaries, along with references to the location of illustrative quotations (i.e., transcript, page, and line numbers), into the appropriate cell in the chart.

**Mapping and interpretation.** Once charting was completed, I reviewed the data across the charts I developed during the previous phase of analysis with a focus on identifying patterns and relationships in order to draw both descriptive and explanatory conclusions clustered around each of the dimensions and sub-dimensions in my conceptual framework. The original research questions to be addressed in this phase of my study guided this process. For example, I used descriptive accounts to illuminate the range and nature of program evaluation activity across Canada’s largest public art museums and explanatory accounts to make sense of any observed variations in these museums’ capacities for evaluation. Moreover, identifying museum practitioners’ evaluation challenges and resource support needs allowed me to later formulate a
number of specific considerations aimed at improving practice and increasing capacity across the sector.

Findings from this phase of my study were used to refine the *Conceptual Framework of Evaluation Capacity in Art Museums* based on the insights I gleaned from the data (see Chapter 6) and were reported under headings that reflected the key dimensions and sub-dimensions in my conceptual framework. Within each of these broad sections, I presented the interview data in a conventional format that interspersed authentic citations in summaries to both assist readers in keeping track of themes (Thody, 2006) and increase the trustworthiness of the research (Patton, 1990).

**Phase Two: Case Studies**

The second phase of my research involved conducting qualitative case studies (Yin, 2014) of two art museums that, based on my interview findings, I identified as operating at the most developed\(^{11}\) level of capacity for evaluation in Canada. This phase in my research was designed to explore, in an in-depth way, what such capacity looked like in the Canadian art museum sector and to better understand the cultural and contextual factors that were at play in shaping it. Included in this focus was an interest in understanding how these museums’ capacities for program evaluation compared to their capacities to conduct studies in other domains of visitor studies inquiry (i.e., scholarly research, market research, and exhibition evaluation). Not unlike the preceding interview study, this phase of my research also sought to further test the extent to which the *Conceptual Framework of Evaluation Capacity in Art Museums* could be used to accurately describe the phenomenon of EC in this setting, especially

\(^{11}\) The term *most developed* capacity refers to the highest level of EC reported by an art museum. It does not necessarily connote exemplary capacity nor is it meant to equate to the highest level of EC reported in another sector (e.g., government).
developed capacity. The research questions that framed this phase of the study were thus both
descriptive and explanatory:

1. What are key dimensions that describe EC in the Canadian art museum sector?
2. How is overall EC manifested in those Canadian art museums most active in evaluating their educational programs, and which factors and conditions, according to practitioners’
   self-reported understandings, enabled their capacities to develop to their current levels?

Overall, the insights gained from this phase of my research were used to (a) further
validate and finalize my conceptual framework, (b) shed empirical light on the phenomenon of
developed EC in selected art museums, and (c) share knowledge aimed at both improving
evaluation practice and increasing capacity for evaluation across the Canadian art museum
sector. In the following sections, I briefly overview case study research, provide a rationale for
selecting a case study research design, and describe how I selected cases, recruited participants,
and collected, analyzed, interpreted, and reported the case study data.

**Overview of and rationale for the use of case study research.** While several texts are
available for conducting case studies (e.g., Creswell, 2007; Merriam, 1998; Stake, 1995, 2006;
Yin, 2003, 2009, 2014), scholars have differing opinions about what defines this type of
research. Stake (1995), for instance, characterizes case study as a unit of analysis (i.e., an issue or
phenomenon occurring in a bounded system). Other conceptualizations present case study as the
process (i.e., a methodology or research strategy) and/or product of systematic inquiry as well as
an object of study (e.g., Creswell, 2007; Denzin & Lincoln, 2005; Merriam, 1998; Stake, 2006;
Yin 2014). Following this latter line of thinking, I also chose to view case study in a multi-
faceted way (i.e., as a unit, process, and product of inquiry).

I selected case study research for this phase of my study because it is impossible to obtain
a true picture of the phenomenon of developed capacity for evaluation without considering the context within which it occurs. What is more, a deeper understanding of this phenomenon can be obtained by studying more than one case. Focusing on two cases therefore allowed me to explore why the art museums selected for case study had the highest levels of EC in the country and how they got that way. This case study design likewise enabled me to (a) test whether EC in the Canadian art museum context could be described using the dimensions in my conceptual framework, (b) explain how the four dimensions conceptualized as influencing factors affected those framed as capacity indicators, and (c) apply the lessons learned from this phase of my study to other art museums interested in building their capacities around evaluation.

In undertaking the present phase of my study, I primarily relied on Yin’s (2014) most recent book on case study research design and methods. In it, Yin (2014) offered a two-fold definition of case study that deals with both the scope of a study and the methodological characteristics that are the features of a case. The first part of his definition states that “a case study is an empirical inquiry that investigates a contemporary phenomenon (the ‘case’) in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident” (p. 16). The second component reads as follows:

A case study inquiry copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result benefits from the prior development of theoretical propositions to guide data collection and analysis. (p. 17)

This two-fold definition illustrates how case study comprises an all-encompassing research method that covers study design as well as data collection and analysis. Below, I summarize the
key tenets of Yin’s (2014) approach to case study research.

**When to use case study.** According to Yin (2014), case study research is ideally suited for examining contemporary events over which an investigator has little or no control and should be considered when a researcher wants to understand a real-life case in depth and believes that such an understanding involves contextual conditions pertinent to the case. Additionally, Yin (2014) suggested that for explanatory investigations (i.e., those asking how and why questions) case study has a distinct advantage.

**Types of case studies.** Besides categorizing case studies as either exploratory, descriptive, or explanatory, Yin (2014) identified two main types of case study designs considered to be variants within the same methodological framework—single or multiple case studies. Within either design, there can be one (i.e., single case holistic, multiple case holistic) or multiple units of analysis (i.e., single case embedded, multiple case embedded). Yin (2014) explained that while a single holistic case allows a researcher to understand one unique, extreme, or critical case, using more than one case enables a researcher to explore similarities and differences within and among cases and to generalize beyond a case. He argued that multiple case study designs follow a replication logic and described how such designs can be used to either predict similar results (literal replication) or contrasting ones (theoretical replication) (Refer to Yin, 2014, for full descriptions).

**The role of conceptual frameworks.** Yin (2014) asserted that the development of a conceptual framework (comprised of theoretical propositions representing key topics from the research literature) at the outset of an inquiry is a necessary component of case study—one that distinguishes this form of inquiry from other related qualitative research methods. Yin (2014) noted that conceptual frameworks serve several purposes. Namely, they help with (a)
determining the direction and scope of a study, (b) guiding data collection and analysis, (c)
structuring a final report, and (d) generalizing the findings from a case study (see also Miles &
Huberman, 1994). Conceptual frameworks, he argued, should continue to be developed as a
study progresses and the relationships between the proposed constructs are corroborated,
modified, rejected, or advanced.

Data sources. Yin (2014) stated that “case study’s strength is its ability to deal with a full
variety of evidence” (p. 12)—a strategy that also enhances data credibility (Patton, 1990). Data
sources may be qualitative or quantitative and can include, but are not limited to, the following:
documents, archival records, interviews, surveys, direct observations, participant-observations,
and physical artifacts.

Analysis. While case study does not claim any specific methods for data analysis and the
methods engaged in will depend on the type of case study selected, Yin (2014) described several
techniques for analysis. These included: pattern matching, linking data to propositions,
explanation building, time-series analysis, logic models, and cross-case synthesis (see Yin, 2014
for full descriptions). Regardless of the technique selected, Yin (2014) stressed that in the
analysis phase a researcher must integrate the various data collected in an attempt to understand
the overall case and answer his or her research questions.

Reporting a case study. Baxter and Jack (2008) noted that “there is no one correct way to
report a case study” (p. 555). When a multiple case study design is selected, a typical reporting
format consists of single case reports followed by the presentation of cross-case results. Other
ways to report case studies suggested by Yin (2014) included: telling the reader a story,
providing a chronological account, following a series of questions and answers, and addressing
the propositions that initially formed a conceptual framework. Regardless of the format chosen, a
researcher will need to report the analytic generalizations or the lessons learned (Lincoln & Guba, 1985) from a case study in their final reporting. Stake (2006) reinforced this notion when he said, “It would be a mistake if a multicase researcher fails to disclose whatever generalizations appear evident from the data, in a tentative way” (p. 90).

**Selection of cases and recruitment of phase two participants.** As mentioned previously, my cases were selected on the basis of one central criterion: they had to be operating at the most developed level of EC in Canada. The selection of cases therefore followed an outlier purposeful sampling strategy (i.e., exemplars of excellence) (Patton, 2015) where information-rich cases are selected because they are “unusual or special in some way, such as outstanding successes” (p. 277). I chose this sampling strategy because it was designed for generalizability in the sense that it aims to yield insights about “principles that might be adapted for application elsewhere” (Patton, 2015, p. 710). I selected two art museums for in-depth case study in view of the fact that, as Yin (2014) pointed out, “even with two cases you have the possibility of direct replication” (p. 64). These museums were the Montreal Museum of Fine Arts (MMFA) and the Art Gallery of Ontario (AGO). Each case was bounded by definition and context (Miles & Huberman, 1994). Thus, within each of these cases lay a singular unit of analysis (i.e., EC as defined in my conceptual framework) embedded within a particular context (i.e., a Canadian art museum). It is important to note that a third Canadian art museum that matched my selection criterion and was therefore selected for case study declined to participate in this phase of the research. No other Canadian art museum—beyond the MMFA, the AGO, and the museum that declined to participate—met the above selection criterion.

Over a period of several months (November 2017 to June 2018), I recruited participants from each the MMFA and the AGO. Recruitment was conducted via a snowball sampling
technique (Patton, 2015) to locate information-rich key informants at each site. Specifically, I began this process by making formal contact with known museum contact persons (i.e., those key informants interviewed during the first phase of this study) and asking them to identify additional staff members who they thought might be good sources of information about visitor studies and/or cultural and contextual considerations at each museum. This strategy was pursued until the chain of recommended informants converged into a small number of core informants with no new names being mentioned. Identified informants included Executive Directors, both Heads of Education and directors of other departments (e.g., communications, membership and visitor services), art museum educators, internal researchers, and others responsible for one of the three domains of visitor studies inquiry (market research, scholarly research, and evaluation) at each museum.

Once identified, key informants were sent, via email, combined letters of information and consent forms (Appendix H) (see Appendix B for an example email script) detailing the purpose of and expectations for participation in the study. Follow-up phone calls were made within a two-week period (see Appendix C for a sample telephone script). Because Yin (2014) argued that “anonymity is not to be considered a desirable choice” (p. 197) in case study, each art museum and its key informants were accurately identified (with their permission). This manner of disclosure ensured that important cultural and contextual information about each case (crucial to my conceptual framework) was not eliminated from this phase of the study.

**Procedures for collecting case study evidence.** Yin (2014) pointed out that “case study’s strength is its ability to deal with a full variety of evidence—documents, artifacts, interviews, and observations” (p. 12). My data collection sources therefore included a mix of Skype and on-site interviews, documents, and direct observations to both offer a comprehensive
picture of EC at each museum and triangulate (Denzin, 1970) the information in the case studies.

**Interviews.** Stake (1995) reminded us that “much of what we cannot observe for ourselves has been or is being observed by others” (p. 64). Since I could not count on being able to directly observe visitor studies underway at the two museums selected for case study, one of my most important sources of case study evidence was the interview. I chose to use an interview guide approach (Patton, 2015) for this phase of my research because the approach makes data collection somewhat systematic for each respondent while also allowing a researcher to decide the sequence of the questions asked. In this way, interviews remain “fairly conversational and situational,” explained Patton (2015, p. 438).

Akin to the development of the semi-structured interview protocol in Phase One of my study, I used the *Conceptual Framework of Evaluation Capacity in Art Museums* to direct the development of the interview guide. Thus, the interview guide grouped the questions (both open and closed-ended) explored during interviews according to the seven dimensions that comprise my conceptual framework. For example, all interview questions asking key informants to report on their museum’s contextual characteristics were listed under the broad category: *Context*. The guide was adjusted depending on the person being interviewed. In Appendix I, sample interview questions are listed per framework dimension and attributed to possible key informants. Using this word table as a basis, I ultimately prepared several different interview guides.

I carried out interviews lasting between 30–60 minutes in length with each key informant recruited from the two art museums selected for case study. Table 4 below lists the names and job titles of the informants from each art museum who agreed to participate in this phase of the study. Key informants were interviewed both at their museum of work to facilitate the collection of relevant documents as well as over the phone. The overall purpose of the interview was
reiterated to informants at the beginning of the interview. During the interview itself, I used the various interview guides developed to focus the discussion on specified topics while also allowing topics of importance to respondents to emerge throughout the process. Once again, I recorded the interviews using a digital recorder, transcribed them in full myself, and edited the transcripts for unity (Holliday, 2002).

Table 4

**List of case study informants per museum**

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jean-Sébastien Bélanger</td>
<td>Head of Membership and Customer Service and the Museum Foundation Office</td>
</tr>
<tr>
<td>Marilyn Lajeunesse</td>
<td>Educational Programmes Officer-Adults and Community Groups</td>
</tr>
<tr>
<td>Nathalie Bondil</td>
<td>Director General, Chief Curator and Curator of European Art</td>
</tr>
<tr>
<td>Pascale Chassé</td>
<td>Director of Communications</td>
</tr>
<tr>
<td>Stephen Legari</td>
<td>Educational Programmes Officer-Art Therapy</td>
</tr>
<tr>
<td>Thomas Bastien</td>
<td>Director of Education and Community Programmes</td>
</tr>
<tr>
<td>Judy Koke</td>
<td>Richard and Elizabeth Currie Chief, Public Programming and Learning</td>
</tr>
<tr>
<td>Shiralee Hudson</td>
<td>Lead Interpretive Planner, Public Programming and Learning</td>
</tr>
<tr>
<td>Keri Ryan</td>
<td>Director, Interpretation and Visitor Research</td>
</tr>
<tr>
<td>Melissa Smith</td>
<td>Coordinator, Gallery Guide Program, Adult Education Officer Program and Access to Art Program</td>
</tr>
<tr>
<td>Nympha Patel</td>
<td>Director, Membership &amp; Database Marketing, Communications &amp; Brand</td>
</tr>
<tr>
<td>Paola Poletto</td>
<td>Manager, Studio and Group Learning</td>
</tr>
<tr>
<td>Sarah Febbraro</td>
<td>Youth Programs Coordinator</td>
</tr>
</tbody>
</table>

**Documents.** Yin (2014) stated that “because of their overall value, documents play an explicit role in any data collection in doing case study research” (p. 107). Recognizing this, I collected documentary information both through internet searches prior to doing fieldwork and at each museum. Examples of documents collected include relevant public domain documentation available on websites (e.g., mission statements, exhibition and educational program descriptions), research and evaluation documents (e.g., program evaluation reports, exhibition evaluation reports) in addition to various administrative documents such as annual reports,
audited financial statements, organizational charts, and strategic plans. Examples of the specific documents selected and the documentary data analyzed per museum are provided in Appendix L.

**Direct observations.** Because “observational evidence is often useful in providing additional info about the topic being studied” (Yin, 2014, p. 114), I also conducted observations at each of the two art museums selected for case study. As I was not actually able to observe any visitor studies underway, my observations focused on other relevant cultural and contextual phenomena of interest to my study such as the availability of educational and community exhibition spaces; the presence, or absence, of educational or interpretive spaces in exhibition galleries; and the size, location, and quality of each museum’s educational facilities. To illustrate, the presence of the Weston Family Learning Centre—a dedicated space for instruction—at the AGO indicated something about the value the gallery places on art museum education. All informal observation notes were recorded in a research journal.

**Case study analysis.** Gale, Heath, Cameron, Rashid, and Redwood (2013) pointed out that the framework approach (Ritchie & Spencer, 1994, 2002) to qualitative content analysis is flexible enough that non-interview data such as segments of document text and field notes can also be analyzed using the systematic procedure. The authors further note that the approach facilitates the comparison of data across cases as well as within individual cases (Gale et al., 2013). I therefore also employed the framework approach to manage, analyze, and interpret the interview, documentary, and observational data I collected during Phase Two. Below, I describe how I used this approach for both single- and cross-case analysis.

**Familiarization.** The analysis process began with transcription. Following this, I re-familiarized myself with my data by re-listening to all the interviews and by conducting a thorough reading of all the case study evidence collected. First impressions of the data and any
thoughts I had about analysis were, once again, recorded in a research journal.

**Coding.** The code list (Appendix J), refined during the first phase of this study, was likewise applied to my entire data set during the present phase of the research. I began the coding process by reading my case study evidence (i.e., transcripts, observational notes, and select document segments) line-by-line and manually assigning the abbreviated sub-codes from my code list to all the data in its textual form. Inductive coding, once more, followed this deductive approach, and any new emergent codes were used to further refine the code list and, by extension, my conceptual framework. Appendix K lists the codes that were applied during Phase Two.

**Charting.** Once coding was complete, I first summarized each coded segment of text and then charted the reduced data into the appropriate matrix coding chart depicted in Appendix G, ensuring, all the while, that the original text was referenced (i.e., with a document, page, and line number) so that its source could be traced. For this phase in my study, the charts were laid out by case (i.e., for each individual case across all the sub-codes falling under a given framework dimension). Yin (2014) argued that charting effectively produces a profile of each case that “permits your analysis to probe whether different cases appear to share similar profiles and deserve to be considered instances (replications) of the same type of general case” (p. 166).

**Mapping and interpretation.** I began this final analytic stage with an examination of the charted data for each art museum on a case-by-case basis. In each separate case, the case study evidence was descriptively analyzed according to my conceptual framework to shed empirical light on the phenomenon of developed EC in the two art museums selected for case study. I subsequently drafted a case report or *portrait* (Lightfoot, 1983) for each art museum which relied on a common report structure to facilitate comparability. Hence, findings for each art museum
were presented under the following headings: introduction and study context, culture, demand for evaluation, resources and supports, capacity to do visitor studies, capacity to use visitor studies, visitor studies activity, evaluation capacity building, and evaluation challenges and needs to align with framework categories. A draft of each case portrait was sent via email to the informants who participated in the case studies for review. A period of two weeks was allotted for this process. Once the portraits had been revised based on informants’ feedback, I provided each art museum with a finalized case portrait for its own internal uses. For example, the considerations presented in each case portrait could help inform the museum’s decision making around evaluation going forward.

Following the within-case analysis and validation exercise, the single-case portraits were used as the basis for comparing similarities and differences between the two art museums selected for case study. Specifically, using an inductive approach, I examined the individual portraits for both common and divergent patterns in search of extrapolations (Cronbach, 1980) and lessons learned that could travel beyond the setting of each specific case. These lessons are presented as a list of recommendations outlining what could be done to strengthen the evaluation capacity of art museums in Canada.

**Strategies for Enhancing Validity and Reliability**

According to Creswell (2009), qualitative validity “is based on determining whether the findings are accurate from the standpoint of the researcher, the participant, or the readers of an account” (p. 191), while qualitative reliability “indicates that the researcher’s approach is consistent across different researchers and different projects” (p. 190). The various strategies I employed to ensure qualitative validity and reliability included: triangulation, member checking,
rich description, a study design that consisted of more than one case, audit trail, as well as critical friend and expert audit review.

**Methodological and data source triangulation.** The information (i.e., interview data from multiple key informants, documents of many kinds, and observational notes describing diverse events and phenomena) I obtained using three different data collection methods (i.e., interview, direct observation, and document review) was brought together and compared during the second phase of my study to illuminate various aspects of developed EC. In this way, both methodological and data source triangulation (Denzin, 1970) were used to strengthen the reliability as well as internal validity of the case studies (Merriam, 1998; Patton, 2015).

**Member checking.** Member checking is viewed by many social science researchers (e.g., Creswell, 2009; Merriam, 1998; Patton, 2015) as an important approach to analytical triangulation. Member checking was used in the second phase of my study to determine the “accuracy, completeness, fairness, and perceived validity” (Patton, 2015, p. 668) of the within-case data analysis by having the key informants in each case studied review and comment on their draft case portraits. In the first phase of my study, interviewees were provided with the opportunity to both review their individual transcripts as well as an initial draft of Phase One results.

**Rich description.** To both add validity to my case studies and enhance the qualitative generalizability of their findings, rich description was used to paint a detailed picture of each of the two art museums selected for case study so that readers might determine how closely their own situations match the research situation and, hence, whether the lessons learned are transferrable to their context (Creswell, 2009; Merriam, 1998).
**Study design.** Yin (2014) explained that using more than one case enables a researcher to explore similarities and differences within and among cases and to generalize beyond a case. As noted above, he stated, “Even with two cases you have the possibility of direct replication” (p. 64). A case study design that used two cases to study the same phenomenon was therefore selected for the second phase of my study to add validity to the qualitative generalizations or lessons learned that were extrapolated from the data.

**Audit trail.** Guba and Lincoln (1981) pointed out that independent judges can audit the findings of a study by following the trail of a researcher. For this to take place, however, they noted that a researcher must clearly describe his or her qualitative research procedures. The five interconnected stages in the framework approach to qualitative content analysis (Ritchie & Spencer, 1994, 2002) explicitly describe the processes that guided the systematic analysis of data from initial management through to the development of descriptive and explanatory accounts. Employing the framework approach to manage, analyze, and interpret the qualitative data collected during both phases of my study therefore allowed me to maintain an effective and transparent audit trail which enhanced the reliability of my study and the qualitative generalizability of its findings (Creswell, 2009; Merriam, 1998).

**Critical friend and expert audit review.** Patton (2015) viewed using critical friend\(^{12}\) review as a form of analyst triangulation. A trusted person (i.e., another graduate student who is familiar with my work) was therefore asked to cross-check my codes to ensure qualitative reliability, comment on my findings as they emerged, and review the entire study at its conclusion. Additionally, my doctoral committee served to further enhance the overall validity of my study by auditing my work to render judgment about the quality of both its process and

---

\(^{12}\) Costa and Kallick (1993) use the term critical friend to define a person who “asks provocative questions, provides data to be examined through another lens, and offers critiques of a person’s work as a friend” (p. 49).
product results (Lincoln & Guba, 1986).

**Research Ethics**

In advance of conducting this research, I sought and obtained ethics clearance from the General Research Ethics Board of Queen’s University (see Appendix A). In addition to this, I distributed combined letters of information and consent (Appendix D and H) to participants which were signed and collected prior to conducting interviews and case studies. These combined documents included written descriptions of (a) the nature of the research, (b) the specific data gathering activities to be carried out, and (c) the expected burden on the participant. Likewise, the combined letters of information and consent also clearly lay out my plans around both confidentiality and the publication of research findings (Stake, 1995). While individual practitioners’ names were kept confidential in the first phase of the study, I disclosed (with their permission) the identities of both the individuals participating in and the names of the art museums selected for case study.
Chapter 4
Phase One Findings

In this chapter, I report the results of my conversations with 13 art museum professionals who agreed to participate in an interview study focusing on their current capacities to evaluate their various educational and public programs. The interview findings are discussed in the following sections, representing a mix of a priori (i.e., the dimensions present in the Conceptual Framework of Evaluation Capacity in Art Museums) and emergent themes. These include: context, culture, demand for evaluation, resources and supports, capacity to do and use program evaluation, program evaluation activities, evaluation capacity building, practice-based challenges and limitations to capacity, practitioner’s self-reported evaluation needs, and reflexivity and opportunities for the future (see Appendix J for the final list of master and sub-codes applied during Phase One).

Context

While I did not ask key informants to directly report on their organizational contexts, reserving the in-depth examination of this influencing factor for the second phase of this study, there were certain contextual characteristics that, nevertheless, emerged in the interview data as bearing on a museum’s overall capacity for evaluation. These sub-dimensions included those already captured as a priori sub-codes (i.e., size, type, location, age, and institutional structure) on my start list of codes (Appendix F) as well as an additional sub-dimension/code that emerged from my analysis of the interview data—namely, an art museum’s extant level of institutional stability versus change.

Size. In general, the museum practitioners I spoke with perceived larger art museums as having more resources, including evaluation resources, to draw upon. Despite this perception,
some interviewees from larger-budget museums lamented that, for the size of their institutions, they were still quite under-resourced. As one museum practitioner noted, “We must have been the only institution in North America who didn’t have a Head of Public Programs.” Although larger operating budgets were generally viewed as advantageous and indicative of greater EC, some practitioners also saw certain benefits to working in art museums with smaller annual gross revenues (i.e., under 10 million). These institutions were perceived as being less formalized in terms of their organizational structures, resulting in more positive and egalitarian working relationships between educational and other museum staff members. A Head of Education explained, “Because we’re a mid-sized gallery . . . we’re able to work so well together . . . and I have a strong say at the management table.” Being able to engage in ongoing reflective practice as a result of practitioners’ close proximity to programs was another reported advantage of working in a smaller institution:

Because all of us are so hands on and so involved directly with our clientele we almost have an hour by hour feedback system on how people are feeling about the program. And we, as a department, meet so often informally that we’re kind of always in that moment of reflection and assessment about what it is that we are doing, how we are doing it, what we need to be doing differently, and so on and so forth. Because we’re so small I think we have that advantage.

**Type.** Perhaps unsurprisingly, Crown corporations were perceived as having a higher capacity for evaluation than those museums not operated by the federal government of Canada. More importantly, however, the museum practitioners I spoke with thought that visitor studies in art museums lagged behind other museum types (e.g., natural history, science)—an observation that has also been made by others (e.g., Hicks, 1996; Luke & Knutson, 2010). As
one Head of Education noted, “Art museums in general are behind the curve when it comes to doing visitor research and evaluation and really using it to inform their exhibition and program development process.” The discrepancy between art and non-art museums’ capacity for evaluation was attributed to several differing factors including the fact that evaluation is not a new function in, for example, science museums. While additional research is needed to describe the extent to which ‘type’ impacts a museum’s capacity for evaluation, this finding suggests that art museums could stand to learn from the evaluation experiences of different types of museums.

**Location.** Numerous practitioners also reported thinking that location or, more specifically the accountability context within which an art museum is situated, impacted its EC, noting that art museums in what they perceived to be higher accountability contexts (i.e., the US and UK) were generally more active in evaluation. One Head of Education observed, “In art museums in the United States evaluation is more embedded in their practice. If you look at the Isabella Stewart Gardner Museum or the Detroit Institute of Arts, I think visitor studies and evaluation are built in to their operations.” In this case, such art museums’ higher capacity levels were associated with significant funder demand for accountability information coupled with concomitant funds for evaluation. Looking to art museums in other countries and accountability contexts could therefore be an important next step in understanding the factors that could support better evaluation practices and greater EC in Canada.

**Age and level of institutional stability.** While age or level of maturity only appeared to influence an art museum’s EC when an institution was newly established (i.e., one museum had only recently opened its exhibition spaces to the public), the interview data revealed that level of institutional stability was a key contextual characteristic negatively influencing organizational capacity for evaluation across Canadian art museums. The majority of the museum practitioners
interviewed reported that their institutions were either directly experiencing or had recently been affected by some manner of organizational change. For example, four interviewees reported that their institutions were either in the process of moving forward with a major renovation or expansion project or raising money to acquire a new facility or expand/renovate an existing one. Three art museums were either engaged in a variety of different organizational development activities (e.g., strategic planning, visioning, mission revision, etc.) or rolling out new operating plans and procedures informed by such change processes. A further six art museums reported having experienced recent turnover within their senior leadership teams. “We’ve seen two directors come and go in the past four years. We now have an interim director who is actually our before-last director, and we’re in the process of hiring, hopefully, a new director,” remarked one Head of Education.

Not only were such organizational changes reported as being disruptive to art museum’s programming (e.g., forcing educational programs off site in the case of large-scale capital projects), they also impacted their evaluation practices and capacities in several different ways. Most notably, both expansion/renovation projects and organizational development activities, while important in their own right, were perceived as siphoning considerable financial and human resources away from evaluation. For example, one Head of Education explained that the more staff time was devoted to strategic planning at her museum, the less time was available for “deeply analyzing what we already do.” Another commented on why a planned renovation project was likely to have adverse effects on her evaluation budget:

I know next year will be a very tough year for us financially because the original building of the museum is being closed down—we’re taking all of the collections out and we’re basically re-doing all of the inside to bring it back to a higher level of quality. So that’s a
huge project that’s draining a lot of resources.

Similarly, turnover in leadership was often reported as resulting in the loss of an art museum’s primary evaluation advocate. When asked why her institution was not engaged in evaluating educational programs to the same degree as in the past, one Head of Education explained, “Our latest director re-wrote my job description and when I got the new job description it didn’t have evaluation in it . . . and so it hasn’t happened. He had other priorities.” This finding aligns with Cousins, Goh, and Clark’s (2006) observation that leaders can have significant influence on evaluation’s enculturation while adding an important nuance. Namely, that organizational leaders can hinder the incorporation of evaluation into organizational culture just as much as they can promote it.

Organizational structure. At least two practitioners expressed being challenged by issues of power embedded in traditional hierarchical organizational structures (see Morgan, 2015)—structures that privilege, what one Head of Education called “the curatorial kind of rule.” She went on to explain the way in which traditional reporting structures prevented her museum from both being more effective when writing grant applications and meeting funders’ accountability requirements:

There is something broken about this framework that we work within—it doesn’t jive with the expectations that are increasing from funders and our publics to be accountable for the resources we’re spending and the work that we do. For example, our head curator is the one who initiates the multi-year grant application processes, but he doesn’t speak the language of engagement. So, there’s a discrepancy between the skills that people in curatorial departments have and those required to develop projects that align with the objectives that funders are laying out. And I’m not consulted about any of this. It’s
difficult to make things connect with people without being involved on the front end.

This finding supports calls made by numerous scholars (e.g., Holman-Conwill, 2003; Murawski, 2018; Samis & Michaelson, 2017) for museums to rethink their internal organizational structures and start including the entire museum staff in a holistic and non-hierarchical manner. Significantly, this contextual sub-dimension was observed to be connected to the cultural sub-dimensions of valuing of museum education and capacity for organizational learning described in greater detail in the section below. In other words, contextual characteristics such as a museum’s organizational structure and mandate were found to reflect certain aspects of its institutional culture. Specifically, the value a museum placed on education as well as the degree to which it demonstrated markers of organizational learning such as clarity and support for mission and vision, shared leadership, and teamwork—markers that are facilitated by flat decentralized organizational structures (Goh, 2000).

**Culture**

The interview questions that enabled investigation of this framework dimension focused on (a) better understanding the value that Canadian art museums placed on both education and evaluation and (b) determining the degree to which these aspects of organizational culture impacted on their evaluative practices and capacities. Considerable evidence was found in the first phase of this study to support both the use of the *culture a priori* master code and the sub-codes of *valuing of museum education* and *openness to evaluation* on my start list of codes (Appendix F), suggesting that both sub-dimensions can contribute to influencing an art museum’s overall EC. Yet a third sub-code emerged from my analysis of the interview data—*learning orientation*. Thus, the extent to which a museum could be classified as a *learning organization* (Senge, 1990) also appeared to either limit or promote capacity. Summarily, not
only did an art museum need to value evaluation to engage in it, but also what it thought was important (e.g., its educational vs. exhibition programs), while sometimes competing, drove what got evaluated.

Valuing of museum education. Roughly half of the art museum practitioners I spoke with reported thinking that education played an important role at their institutions. An art museum was characterized as valuing education when it placed learning and engagement at the forefront of mission statements, strategic plans, and marketing campaigns, and supported the work of education departments by providing them with adequate resources (e.g., funding, staff). In such museums, education also often held prominence within institutional structures such as when Heads of Education comprised senior management teams. The traditional hierarchical divide between curatorial and educational efforts in these museums was likewise reported as being minimized. For instance, practitioners perceived education as being highly valued by their institutions when they were involved in exhibition development and interpretation, invited to work on socially engaged curatorial projects, or given license to host educational projects within exhibition spaces generally reserved for professional artists. One Head of Education described the working relationship between education and curatorial staff at her art museum in the following way:

In many museums curatorial staff and education staff won’t necessarily work together—curatorial does an exhibition and then they pass the baton to education. With [art museum] it’s different. The managers in curatorial and education are increasing working collaboratively. Before we actually decide on exhibitions, we have discussions about how we are going to work together.
In general, such museums reportedly dedicated more time to evaluating programmatic impacts, employed staff with higher levels of evaluative knowledge and skills, and were more likely to engage in evaluation capacity building (ECB) than those art museums that did not place an equal emphasis on learning. Another four practitioners reported feeling that education was either valued or becoming increasingly valued at their places of work, but that it was still marginalized in comparison to curatorial activities. While some of these practitioners accepted the marginalization of educational endeavours as a matter of course in art museums, others expressed thinking that the prioritization of exhibitions over educational programming had negative consequences not only for program implementation, but also evaluation. As one practitioner noted:

Our resources go towards evaluating the success of our ticketed exhibitions because they bring in the most people. I don’t need to know that much more about those exhibitions at this point, but I can’t not evaluate them. There are programs that I think should be evaluated because they’re big elements of the [gallery’s] operations, but we don’t think of them that way because they operate at arms-length from the exhibitions.

The remaining two art museum practitioners I spoke with wanted to see education take a much more prominent and integrated role in their institutions. These practitioners expressed being challenged by scarce resources, power imbalances, and staffing structures that limited their abilities to collaborate with colleagues at other levels of the institution (see organizational structure above). Needless to say, such challenges were reported as hindering these practitioners’ abilities to conduct program evaluations and use their results for varied purposes.

**Openness to evaluation.** All the art museum practitioners I spoke with recognized both the need for and value of evaluation. “It is absolutely integral to what we do,” said one
interviewee. Another went as far as to say, “If museums are to survive long-term, they need to be able to articulate why they matter in our communities . . . and I think [visitor research and evaluation] is the path to getting there.” Despite Canadian art museum educators’ openness to evaluation, the results of the first phase of this study revealed that evaluation as a value was not necessarily evenly integrated across all departments or levels of any one museum. The degree to which an art museum possessed an overall climate that was supportive of evaluation also varied across the sector.

Only two of the practitioners I spoke with reported that their museums had successfully infused evaluative thinking into their organizational cultures as part of their overall commitments to visitor learning (see Chapter 5). Both art museums demonstrated the following characteristics (a) leadership demand for and support of evaluation, (b) ongoing pan-institutional engagement in a wide range of research and evaluation activities as part of mission fulfillment, and (c) significant capacity to both do and use evaluation. In terms of the latter characteristic, these art museums either employed staff whose responsibility it was to manage the production of evaluative knowledge both at multiple levels of the institution and across various departments and programs or conducted evaluations in collaboration with numerous outside experts working in educational, community, health, and other settings. The results of such efforts were described as not only being used for instrumental purposes (e.g., decision making and reporting), but also, and perhaps more importantly, to contribute to knowledge and promote advances in fields such as museum education, the health sciences, and the creative arts therapies.

Another two practitioners reported that they were beginning to see a positive shift in how their museums think about evaluation—from an occasional nicety to a necessary ongoing activity. “I would say that there is a growing interest in and a growing acceptance of evaluation,”
said one Head of Education. Nevertheless, these practitioners also noted that evaluation was not yet equally appreciated across all departments. As another practitioner explained, “It’s really education and marketing that have the most buy-in into the usefulness of visitor studies and evaluation because we’re always engaged in collecting data.”

The remaining practitioners I spoke with reported thinking that the art museums at which they worked had yet to realize the value and utility of evaluation. A lack of interest in or awareness about evaluation or its importance in addition to either negative perceptions of or outright resistance to evaluation on the part of groups or individuals were perceived as contributing to evaluation’s slow uptake in these institutions. “Directors they don’t really care. I think evaluation’s just never been that much of a concern,” said one Head of Education referring to her senior management team. Another practitioner noted, “I think, in general, people don’t love [evaluation] . . . it’s always dry and time consuming. I think that’s sort of felt across the institution.” Yet another explained, “The first year there was an expectation that everybody from each department spent time in the gallery and did these interviews, but that did not last very long. There was a lot of push back.”

**Capacity for organizational learning.** The term organizational learning is widely understood to refer to the capacity of an organization to modify its practices based on the knowledge it acquires (Garvin, 1993; Senge, 1990). Goh (2000) argued that organizations that possess this capacity (i.e., learning organizations) share certain key qualities. Namely, they tend to, among other things, promote shared leadership, emphasize teamwork, and foster a climate of continuous improvement and experimentation. The concept of organizational learning is increasingly becoming part of the conversation in the ECB literature. To date, however, it has
mainly been positioned as either a driver towards (e.g., Preskill & Boyle, 2008) or a desired outcome of (e.g., Cousins, Goh, Clark, & Lee, 2004) ECB.

The findings of this research go further to suggest that the existing learning capacity of an art museum influences its openness to evaluation, which ultimately carries implications for its evaluative practices and capacities. Specifically, several practitioners attributed a lack of institutional commitment to evaluation to their museum’s broader lack of commitment to organizational learning. Rather than being committed to improvement and experimentation, these museums were described as resistant to both risk and change. “We’re a very risk-adverse institution,” noted one Head of Education. Another commented, “Management would need to be willing to look at the changes visitors are asking us to make and that’s a huge thing, that’s a big expectation for an institution that has a hard time shifting.” This finding demonstrates that both ECB and evaluation mainstreaming may be best advanced in those art museums with an already high capacity for learning.

**Demand for Evaluation**

The interview questions associated with this framework dimension aimed to ascertain the drivers of evaluation in the Canadian art museum sector and understand the types of evaluation information practitioners were required to collect to meet both internal and external accountability demands. The findings from the first phase of this research supported the use of the demand for evaluation a priori master code and the sub-codes of internal impetus and external pressure on my start list of codes (Appendix F). The results likewise indicated that while many funders require the art museums they support to conduct some form of evaluation, the information that practitioners collect for this purpose is information that they are already collecting for either internal accountability or learning purposes. Moreover, most practitioners
are asked to communicate evaluation results to both funders and their museum’s leadership using output measures, not outcome measures.

**Internal impetus.** Five of the art museum practitioners I spoke with reported that they conducted evaluations, among other reasons, because they were required to by their institutions. Referring to why he undertook his last program evaluation, one Head of Education said, “Senior management wanted the evaluation done.” Most of the remaining practitioners interviewed reported that while they were not formally required to evaluate their educational programs, there was an unspoken expectation to do so. “It’s never been asked of us, but I know if all of a sudden there was no evaluation happening there would be questions asked,” noted one Head of Education. Another commented:

> Evaluation is essential . . . I think [the museum leadership] understands that. To what extent they put pressure on us to do formal evaluations, I would say I don’t feel the pressure that way, but I think they just assume that we’re doing it.

Where the internal demand for evaluation was made explicit, the overwhelming majority of practitioners said that their institutions expected to be provided with information about program outputs (e.g., the number of programs offered, number of participants served, amount of revenue earned, descriptions of activities, participant satisfaction) rather than outcomes or impacts. As one Head of Education noted, “I don’t have any pressure to or expectation on behalf of the management team to collect impact data.” When asked why his art museum tended to measure program success based on outputs, another practitioner explained, “It’s really important to measure attendance and revenue for any institution, and for us as well it’s really important because a strategic priority at the gallery is to increase attendance.” While measuring such outputs is undoubtedly necessary for art museums to ascertain if they are achieving certain
business goals, this finding also suggests that these institutions may not fully understand the value of outcome measures over output measures, pointing toward a need for greater evaluative knowledge within the sector.

**External pressure.** The results of the first phase of this study indicated that the evaluation expectations of funders vary not only from funder to funder, but also from program to program. “Different funders have different expectations,” said one art museum practitioner. While there were exceptions to this rule, the majority of the practitioners I spoke with reported that their funders, not unlike their institutions, wanted to be provided with output information. As one Head of Education noted, “As long as you mention how many people came to the event, give a description of the event, demonstrate that it was appealing, and give the funder visibility, I mean that’s all they’re really interested in.”

Further to this, interviewees reported that their funders did not have specific expectations about how art museums should undertake their evaluations or who should be responsible for conducting them. No funder expected them to use an external evaluator. For these reasons, practitioners perceived the evaluation expectations of funders as relatively easy to meet. “I don’t find it challenging,” noted one Head of Education. Three of the art museum practitioners I spoke with said that some of their funders expected them to report on the outcomes of their programs. Describing one such funder’s expectations, a Head of Education said:

> There is one funder in particular who asks you to, almost like a school curriculum rubric, predict outcomes and say how you are planning to meet those outcomes and, on the reporting side, what your outcomes were and how you met them.

Although such funders were reported as being rare, one Head of Education described thinking that funders’ interests in obtaining information about program outcomes were on the rise. She
noted, “I think we’ve noticed that there is a shift in how funders think or would like to think of their impact you know? . . . Demonstrating impact on communities served is becoming very important.”

The above findings align with those of other researchers (e.g., Adams, 2012) who have noted that many museum practitioners say that their supervisors, administrators, and/or funders are only interested in output data. They likewise suggest that as much as work needs to be done to help art museums more effectively measure programmatic success, more research is needed into Canadian funders’ evaluation expectations to determine why so many of them appear to expect art museums to report on their programs using output measures alone.

**Resources and Supports**

The interview questions that enabled investigation of this framework dimension focused on revealing the type and extent of the resources and supports that were available for program evaluation in the Canadian art museum sector. The findings of the first phase of this study supported the use of the *resources and supports a priory* master code and the sub-codes of *time, human resources, financial resources, technology,* and *other supports* on my start list of codes (Appendix F). Overall, the Phase One results revealed that there is a lot of room for improvement in the area of funding and other supports for evaluation in such museums. As such, it is unsurprising that practitioners identified a lack of evaluation resources and supports as both a major challenge and a pressing need (see the *Practice-Based Challenges and Limitations to Capacity and the Practitioners’ Self-Reported Evaluation Needs* sections).

**Time and human resources.** While the majority of the art museum practitioners I spoke with reported having primary responsibility for the evaluation of their educational programs, they received varying degrees of support with data collection, analysis, and reporting from staff
members working both within and outside of their departments (e.g., communications, membership and visitor services, development). “I’m very grateful that I have an assistant who crunches the numbers for me, and then I can look at it and make decisions,” said one Head of Education noted. Another noted:

The reporting is a joint effort between education and development who are in close contact with our funders, and we provide them with the information we’ve collected and work with them on the report to make sure it’s accurate and deliverable.

These practitioners reported dedicating roughly 10 percent of their time to evaluation activities. The one exception to this rule was a practitioner who was employed by her museum in a permanent research capacity. She reported spending 20 percent of her time on evaluation and was assisted in her work by an additional five staff members. A total of five interviewees further reported having enlisted various external experts (e.g., university researchers, graduate students, professional evaluators, marketing consultants) to provide additional program evaluation services. Such services were largely episodic in nature. One practitioner, however, reported that her museum operated on a partnership model of evaluation where studies were undertaken by external experts working in concert with managers across various programs. In this case, evaluation knowledge was produced on an ongoing basis, and the time devoted to this activity increased significantly due to the involvement of multiple stakeholders.

Financial resources and other supports. The art museum practitioners that I spoke with reported that they did not receive any financial resources for conducting evaluations from their institutions. Such job responsibilities were already covered by their salaries. As one Head of Education noted, “There’s no special evaluation fund.” When asked about the financial support they received from funders, only two practitioners reported that their funders either provided
dedicated monies for evaluation or allowed a portion of their project funds to be used for evaluation purposes. “I think it can be up to 10% of the grant total that can be dedicated specifically to evaluation,” explained one education staff member.

The findings of the first phase of this research also revealed that practitioners likewise received limited non-financial supports (e.g., advice on evaluation, technical assistance, training support, or evaluation tools/resources) from both their institutions and funders alike. The most frequently mentioned form of evaluation support available to them was training support. For example, two practitioners said that their art museums made funds available for them to participate in direct ECB activities. Another practitioner reported receiving evaluation training directly from a funder. She noted:

Every year the (foundation) runs an information seminar that is largely directed at evaluation. So, they want to help you figure out how to write your grant, but also how to report on it, and a lot of their advice is transferrable.

Because such training provisions were the exception rather than the norm, this finding matches the observations of Hall, Phillips, Meillat, and Pickering (2003) who assert that many funders who demand evaluation do so without recognizing the need to provide financial support or to work with organizations to build their evaluative capacity and expertise. With respect to the availability of technology, however, the majority of practitioners I spoke with reported having access to computerized data bases that assisted them in compiling, analyzing, and managing evaluation information. Such databases integrated multiple streams of administrative data across various departments on a continuous basis. One museum staff member described how she used such a data base in her work:

Our new point of sale system has a built-in report compiler which we can use to retrieve
data. We have now 100 000 members, and all their data is there. Every teacher that has ever reserved a school visit their information is in there—the times, the dates, the number of students, the number people that have come with them, everything is there.

**Capacity to Do and Use Program Evaluation**

The interview questions associated this framework dimension sought to (a) ascertain the degree to which Canadian art museum practitioners possessed the requisite knowledge and skills to evaluate their educational programs (capacity to do evaluation) and (b) understand their abilities to both make use of evaluation findings and benefit from engagement in evaluation processes (capacity to use evaluation). The results of the first phase of this research supported the use of the *capacity to do and use a priori* master code and both the *knowledge and skills* and *process use* sub-codes on my start list of codes (Appendix F). The original *use of findings* sub-code was further parsed out into *instrumental use, conceptual use, persuasive use, and non-use*. Overall the results indicated that, notwithstanding a few notable exceptions, most practitioners lacked the hard skills to undertake systematic data-based inquiries. They nonetheless possessed sufficient capacity to conduct informal, mostly output-based formative evaluations of their programs (see the *Program Evaluation Activity* section) and use the results of such investigations for a wide range of purposes. A number of process use effects were also reported among those art museum practitioners who had participated in evaluator/researcher-led studies.

**Knowledge and skills.** The interview data revealed that most of the museum practitioners participating in the first phase of this study received little or no formal training in evaluation as part of their professional preparation—a finding that corroborates the results of previous research into the qualifications of art museum educators (Chen-Cooper, 2007; Ebitz, 2005). While their professional preparation varied considerably, the majority of the informants I
spoke with came to the museum profession from disciplines (i.e., studio art, art history, education, art education) where program evaluation, and in some cases social science research methods, are rarely addressed in course work. It is therefore perhaps unsurprising that, with few exceptions, these practitioners reported feeling less than comfortable with the idea of conducting systematic data-based inquiries into their educational programs. This included feeling that they lacked key knowledge and skills pertaining to instrument development and data collection, analysis, interpretation, and reporting. As one Head of Education stated:

I don’t have a background in terms of data collection so it is not my strong suit.

Sometimes I’d be looking at these . . . pie charts you know that Survey Monkey would spit out and trying to analyze the data, and that was a real struggle for me really. I was ready to put [the findings] into practice, but it was the interpretation part that I struggled with.

Such practitioners reported either drawing on their knowledge of classroom assessment when designing data collection instruments, for example, or learning various informal visitor studies and evaluation techniques on the job. Those practitioners without formal training in evaluation who, nonetheless, said that collecting, analyzing and interpreting data, and reporting findings did not pose a challenge for them either (a) had participated in some form of in-service evaluation training in the past (see the ECB section) or (b) were measuring such things as participant satisfaction or number of visitors served (information that is relatively easy for museums to collect) and reporting findings very informally. The following statement helps to explain why one Head of Education did not appear to find program evaluation difficult to conduct:
My assistant collects the attendance data for me . . . and then I read people’s comments . . . So, I don’t find it difficult . . . I feel really comfortable with that. As far as reporting goes, I have no problem saying, “You know that didn’t go as well as I thought . . . and next time I’m going to do things differently.”

While they were in the minority, several art museum practitioners were in possession of both formal evaluation training and significant practical visitor studies and evaluation experience. More specifically, one practitioner held a Master of Arts degree in Educational Psychology with a specialization in Program Evaluation. Another two held graduate degrees in Museology where they took courses in visitor research. One of these two practitioners had worked as both a researcher at a pre-eminent free-choice learning and development organization and a manager of visitor research at a science museum. The other had worked at one of the world’s largest cultural planning and management firms and was currently employed in an internal visitor research capacity—a role she had held since 2011. Yet another art museum practitioner held a Master of Arts degree in Leadership and Training and had significant experience in the field of innovation consulting where she was involved in working on a number of large-scale evaluation projects, including a concept mapping study conducted for a leading U.S. national public health institute. Despite her considerable evaluative knowledge and skills, this latter practitioner, who worked at a smaller-budget institution, reported that she held very little responsibility for evaluation at the art museum at which she worked. She explained why this was the case:

I have a real huge academic and practical background in evaluation so we could very easily set up a very comprehensive formal evaluation program, but I’m really not making
use of hardly any of it because I’m part time and [my director] wants me to be hands-on with my programming so I haven’t drawn on my evaluation experience in the past.

Regardless of their level of knowledge of social science research and evaluation methods, the museum practitioners I spoke with identified their abilities to both understand community needs and how learning occurs in informal settings as well as their interpersonal skills as principal strengths that they brought to the evaluation of programs. Some believed that, because of these strengths and with adequate training and support, museum educators were in the best position to evaluate educational programs in art museums. One Head of Education spoke to this when she said:

I feel like educators are the most poised to do this type of work because just personality wise, mindset wise, and the perspective that they have in terms of having to be a little bit more fluid, more responsive, and they already have that concept of listening in their fibres in the way that they do the work that they do. So that’s where [evaluation] falls very naturally, I think, is in education or public programming departments.

It is important to note that several Canadian art museums, including those that employed staff with the capacity to operate as evaluators, relied on partnerships with outside experts to bolster their own internal capacities for evaluation (see the resources and supports section).

**Use of evaluation findings.** The art museum practitioners I spoke with described the evaluation of educational programs as serving both managerial and accountability functions. Namely, practitioners reported being engaged in a continuous cycle of program monitoring and tailoring while also preparing periodic summative reports for funders and higher-level management. Unsurprisingly, these parallel evaluation purposes were described as leading to a wide range of *instrumental, conceptual, and persuasive* uses of evaluation findings—use
distinctions that pervade in the evaluation utilization literature (e.g., Alkin, 2005; Cousins & Shulha, 2006; Patton, 2008). *Non-use* was likewise noted in several practitioners’ accounts.

**Instrumental use.** Patton (2008) describes instrumental use as the intended use of evaluation knowledge for problem-solving or decision making purposes. Such actions can be improvement- or judgment-oriented depending on whether an evaluation is formative or summative. In the first phase of this study, instrumental use was particularly noted in terms of program improvement. For example, one Head of Education described using evaluation findings to restructure camp programs:

One of the bigger changes that we have made recently is to move our March Break program from being a half-day to a full-day event only. That involved looking at how, over the last couple of years, people have been participating in the programs.

An education staff member with responsibility for evaluation commented on how she used information to make changes to the teacher resources she offered, “We’ve gotten feedback that the teacher preparation materials need to be adapted to the primary grades so we’ve taken that into consideration when designing future study guides.” Another found evaluation findings useful for making staffing decisions, “I remember in the teen art group, rather than having a new instructor every month, they really wanted to have one instructor continue working with them over a period of several months. So, I changed that.” An internal researcher further reported directing findings towards solving an implementation-related problem:

Evaluation helped us understand that people didn’t like waiting a long time to get drinks at the bar. Now we kind of knew that anecdotally, but it really came out in the surveys so we changed the entire set up for the bar, and the next time we surveyed participants that was no longer an issue.
In addition to the instrumental use of formative evaluation, practitioners also described using evaluation findings to make summative judgments about a program’s overall merit and worth, often at the request of a funder. One Head of Education described using evaluation findings to support judgments about a pilot project’s value so that a decision could be made about the continuation of similar programming:

The education department has done a number of learning spaces within special exhibitions, but this was the first time we had such a large foot print and such a high profile in the gallery, and so we really wanted to demonstrate the value of the hub and we also wanted to know did we reach the objectives that we set for ourselves because, for us the hub was quite significant and the cost to create and install the hub was large so we thought we really needed to learn from this experience to decide down the road if we want to do another similar learning space or something more permanent.

Most practitioners described having the autonomy to use evaluation results to make necessary decisions about the programs under their purview. As one education staff member noted, “In terms of smaller program modifications . . . we are able to make those changes internally very easily.” Frustration was nonetheless expressed by a handful of practitioners who felt that they were unable to affect change beyond their own departments. “The biggest challenge is getting the results to move out beyond the project and to have influence more broadly in the institution and in successive projects,” explained one Head of Education.

**Conceptual use.** The term *conceptual use* refers to instances where evaluation findings are used to increase understandings that do not necessarily lead to any direct actions (Patton, 2008). Included in this definition are both evaluation and research efforts that seek to contribute to the generation of new knowledge (e.g., cluster evaluation and basic research) (Patton, 2008).
Several of the art museum practitioners I spoke with noted the conceptual use of findings in terms of both better understanding their target audiences and programming models. For example, one Head of Education explained how evaluation findings helped her gain new insights into the needs of dialysis patients and how to more effectively work with such audiences in future:

We work with a lot of health audiences, but with this one in particular there were just things that we didn’t know about like the fact that you can’t even move without setting off the sensors connected to your body and things like that. So, we learned to do more consultations with the patients and their health care providers before sort of jumping into this type of programming.

An education staff member with responsibility for evaluation meanwhile described how the findings of an evaluation allowed her to gain conceptual insights into how a successful program model could be rolled out across a broader range of partnership programing:

From the project’s inception, with [hospital] we talked about having staff from our two teams share their expertise with each other and so one of the first things we did was train the other side of the team. So, we of course received a lot of training in understanding dementia from them, and then we presented a series of training workshops for the hospital staff in the basics of interpreting art so that when they were in the galleries working on the project they had some understanding of what we were doing. So, I think this upfront two-way training in a partnership scenario is a really good model to work with when implementing partnership programs and we’ve taken this program model and applied it to everything from special education programs for schools to looking at working with young offenders.

In addition to using evaluation findings to gain insights for future program planning, a
small number of art museum practitioners also reported having engaged in some form of knowledge-generating evaluation (Patton, 2008) and using the findings of such studies to identify and share lessons learned about effective practices and programs that might be worthy of application to new settings. “We have published one journal article outlining those results, and another is pending,” said one practitioner referring to recent a co-authored paper in a refereed publication dedicated to innovative aging research. Another two practitioners had recently co-authored a chapter in a book devoted to visitor-centred practices in art museum education and curation.

**Persuasive use.** While in the minority, several of the art museum practitioners I spoke with talked about using evaluation findings to not only increase the status of their educational departments by promoting their work within the hierarchy of an art museum, but also to educate their leadership about the value and utility of evaluation. One Head of Education spoke to the latter point when he said, “There is a bit of an advocacy angle that goes into [evaluation]—to demonstrate the value of programs to visitors and to demonstrate the value of evaluation to senior management.” Yet another educational staff member described using evaluation findings to demonstrate program relevance and make the case for its continued support:

Knowing that the funding from [foundation] was coming to an end, we wanted to gather the information first of all to show them that it was a program that was very well received by the public, and at the same time we needed to gather the information in order to show future funders that “yes, this program really must continue because so many people now have been using it and have come to depend on it.”

**Non-use.** The problem of non-use of evaluation findings was likewise observed across several interviews. “I think that there can be a tendency [in art museums] like in a lot of places to
gather a lot of information, but not really look at it or analyze it or act on it,” one Head of Education observed. In addition to not utilizing information collected as a matter of routine, the art museum practitioners I spoke with described two other variations on non-use that represented some type of ‘failure’ in the evaluation process. As the following quotation illustrates, the first type occurred in instances when development departments drew on museum data bases to prepare funding reports on programs that education staff members themselves did not actually use, “Any report that’s done for a funder . . . I suppose I’d have access to it, but I’m just doing the program.”

The second type of non-use occurred when art museums, in failing to make use of existing evaluation findings, repeated past mistakes. One Head of Education described how this type of non-use can not only result in wasted time and money, but also alienate target audiences: We had this exhibition . . . it was a classic show that we put so much effort and money into it, and it just didn’t perform well. And at the end we were like, “why didn’t people respond to this?” Sometimes we don’t learn from our mistakes because if we would have looked at a survey from five years before that asked our membership about the exhibitions they wanted, classical exhibitions were at the bottom of the list. That wasn’t on the curator or director’s radar at all, and it should have been.

While the presence of non-use permeates the evaluation utilization literature (e.g., Patton, 2008), more research is needed to determine how prevalent non-use is within the art museum sector writ-large.

**The impact of evaluation processes.** Three art museum practitioners reported having worked alongside an external evaluator or researcher on a formal study. While only in one case was the evaluation process used to intentionally build capacity, all three practitioners reported
both learning to think more empirically and acquiring practical evaluation skills as a result of participating in an evaluation—outcomes that reflect two of the six uses of evaluation processes identified by Patton (2008) (i.e., *infusing evaluative thinking into an organization’s culture* and *supporting engagement, self-determination, and ownership*). One education staff member with responsibility for evaluation described the ways in which she was impacted by collaborating with a team of health researchers to evaluate a special program aimed at dementia patients and disseminate the study’s findings in a peer-reviewed journal:

My work with [hospital] has been a really valuable learning experience, and it became capacity building in ways that I had not originally anticipated . . . From working with [hospital] has come an awareness of the kinds of research tools and questions that you look for so that’s something I have applied to looking at other projects as well. So, thinking about identifying research questions and doing reviews of the literature is something that I try and build into new program proposals. Working with the team on the journal article was an interesting process to participate in, too. I learned to look at quantitative data and the importance of not only looking at what the numbers tell you, but what the lack of numbers tell you. I would love to learn a lot more, but I certainly am more comfortable with understanding that kind of approach to research.

In line with an established body of literature (e.g., Patton, 2008), process uses such as those described above had a lasting impact on those art museum practitioners who had participated in an evaluator/researcher-led study. These practitioners maintained strong commitments to data-based decision making, supported the development of evaluative thinking among colleagues, and went on to conduct subsequent evaluations independently. This appeared to hold true even in cases where practitioners were involved in stand-alone studies, regardless of
how long ago those studies took place (e.g., 10 years). This finding suggests that evaluation approaches that are characterized by the engagement of participants in evaluation processes (e.g., participatory, collaborative, and empowerment evaluation) are a promising way of building the internal capacity of art museum practitioners to both do and use evaluation.

**Program Evaluation Activities**

The interview questions associated with this framework dimension were concerned with capturing the nature and extent of the program evaluation activity occurring in Canadian art museums. The results of the first phase of this study support the use of the *evaluation practice a priori* master code as well as the six sub-codes grouped under it (i.e., *type of evaluation; what gets evaluated; data collection, analysis, interpretation, and reporting; type of information collected; who was involved; and satisfaction with the frequency and quality of program evaluation activities*) on my start list of codes (Appendix F). Overall, the results indicate that most practitioners conducted mostly informal formative studies as part of a continuous cycle of improvement and that, in general, practitioners with higher levels of evaluative knowledge and skills in addition to those that either contracted out or partnered with external experts were more likely to not only use more formal data collection, analysis, interpretation, and reporting methods, but also to conduct both summative studies and outcome evaluations.

**Type of evaluation engaged in based on evaluation purpose.** The art museum practitioners I spoke with indicated that, first and foremost, they conducted improvement-oriented formative evaluations that focused on determining both the strengths and weaknesses of programs in order to modify them for the better. As such, the most frequently reported question that focused such investigations was: What do participants like, dislike, or want to change about the program? Formative evaluation was used to prototype activities, evaluate new initiatives, and
monitor the quality of ongoing programs. In all but one case formative studies were carried out by internal program staff as part of their regular activities. One practitioner who worked at a larger-budget museum reported enlisting the services of an external consultant to formally evaluate a popular family program and provide staff with recommendations for its reinvention.

A smaller number of participants (i.e., four) reported undertaking judgment-oriented summative evaluations aimed at determining the overall effectiveness of a program. The questions that guided these types of studies included: Did the program attain its objectives? What impacts were attributable to the program? Should the program be continued, expanded, changed, or discontinued? Summative evaluations were used most frequently to make decisions about the continuation of a pilot project and were carried out by large- and small-budget art museums alike. In contrast to formative evaluations, most summative studies were not only conducted at the request of funders or an institutions’ leadership, but also tended to rely on assistance from one or more external experts. For instance, one practitioner who worked at a medium-budget institution reported partnering with a team of university- and hospital-based researchers and other experts (e.g., post-doctoral researchers, medical doctors, occupational therapists) to evaluate a pilot project aimed at dementia patients and their caregivers. In this case, the summative evaluation was conducted for the benefit of two separate funders.

The one Head of Education who reported leading a summative evaluation internally did so at the request of his Board of Directors. This practitioner worked at a larger-budget museum, and both he and his program staff had received formal training in evaluation (see the Evaluation Capacity Building section) and been engaged in evaluator-led participatory evaluations in the past. As a result, he explained that, “as a department we had internal expertise . . . so, we didn’t have to engage an external firm.” Similar to other research conducted in this area (e.g., Adams,
2012), the above findings suggest that formative studies are both easier and more realistic for art museum practitioners to do themselves. This said, the results of the first phase of this study also indicate that both direct ECB activities and partnerships with external experts can increase the capacity of practitioners to conduct summative studies.

**What gets evaluated.** When asked which programs they evaluated, four of the art museum practitioners I spoke with reported evaluating all or most of their programs. “We’ve managed to pretty much evaluate every single program,” said one Head of Education. Three of these practitioners worked in smaller-budget art museums while the other was employed at a large-budget museum with both an established evaluation culture and a high capacity for evaluation. The remainder of the practitioners I spoke with described evaluating certain core programs on an ongoing basis as a matter of course while also performing periodic evaluations of other programming either in response to external pressure or to address specific issues.

In general, practitioners reported evaluating school programs most frequently. Additionally, they mentioned prioritizing the evaluation of pilot projects and new programs as well as larger, more long-term, and higher-profile programs. “Usually it’s the bigger things like school and camp programs that I look at the most closely,” said one Head of Education. Another one noted, “We have a program called (name) that specifically targets millennials . . . and we wanted to do some evaluation on it because it was a brand-new program.” Practitioners likewise said that they would conduct an evaluation if they were thinking of either revising or discontinuing a program or in the event that a program was unsuccessful. Overall, practitioners reported being less likely to evaluate programs that they described as tried and true. As one practitioner noted, “I think when a program is kind of chugging along and it’s doing well, we don’t usually evaluate it that much.”
While not the focus of this phase of the research, a number of interviewees further reported that they and/or other staff members at their institutions also conducted exhibition evaluation and market research to both determine the effectiveness of exhibits and better understand visitors’ perceptions, demographics, leisure habits, and visiting patterns. These domains of visitor studies inquiry (i.e., exhibition evaluation and market research) will be examined in greater depth in Chapter 5 where I compare program evaluation activity to the full range of visitor research activity occurring within those Canadian art museums selected for case study.

**Data collection, analysis, interpretation, and reporting.** When asked about their data collection methods, the art museum practitioners I spoke with reported using document reviews (e.g., of visitor comments, attendance records, and revenue records) and questionnaires most frequently to gather the data necessary to make decisions about programs followed by informal conversations and unstructured observations. While reported less frequently, participants also described relying on such data collection methods as semi-structured interviews, focus groups, card sorts, artifact reviews (e.g., of student work made following a museum visit), and structured observations such as timing and tracking. Table 5 provides a breakdown of the frequency of use of each method mentioned. While most practitioners reported using more than one method to collect data, in general, practitioners with a higher level of knowledge and skill in evaluation in addition to those that partnered with external experts were more likely to use more formal data collection methods, often in combination with informal approaches. For example, a

---

13 A card sort is a type of survey interview used in informal learning settings that requires participants to place cards containing different scenarios on a Likert-type scale (Grack Nelson & Cohn, 2015).

14 Timing and tracking is a specific type of structured observation used in the museum field that relies on a checklist or protocol to record where people stop, how long they stop, and what they do when they stop during a visit (Yalowitz & Bronnenkant, 2009).
Head of Education who reported having participated in direct ECB in the past (see the ECB section below) described using the following methods to measure the overall effectiveness of a new interpretive space (a) document review of attendance records to determine how many people visited the space, (b) structured observations to evaluate visitors’ movements through and behaviours within the space, and (c) interviews that solicited responses from visitors about their experiences with the various learning activities that comprised the space.

Table 5

*Overview of data collection methods used*

<table>
<thead>
<tr>
<th>Method</th>
<th>Frequency of Reported Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document review</td>
<td>12</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>8</td>
</tr>
<tr>
<td>Informal conversation</td>
<td>5</td>
</tr>
<tr>
<td>Unstructured observation</td>
<td>4</td>
</tr>
<tr>
<td>Structured observation</td>
<td>3</td>
</tr>
<tr>
<td>Interview</td>
<td>3</td>
</tr>
<tr>
<td>Focus group</td>
<td>1</td>
</tr>
<tr>
<td>Card sort</td>
<td>1</td>
</tr>
<tr>
<td>Artifact review</td>
<td>1</td>
</tr>
</tbody>
</table>

The first phase of this study generated moderate data with respect to how practitioners actually analyzed the information that they collected and even less about the methods they used to interpret it. Five of the art museum practitioners I spoke with reported using staff meetings to discuss and make meaning around data. “I’ll get together with my team every two weeks, and we discuss the programs and how they’re going,” noted one Head of Education. Another four practitioners reported using the data analysis and visualization functions embedded in their internal computerized data bases or within software such as Microsoft Excel, Survey Monkey, or Tableau to calculate descriptive statistics for their data sets. One practitioner who had partnered with external experts on an evaluation reported using qualitative content analysis to analyze data.
collected using a questionnaire and structured observation protocol. Two practitioners did not know how the data they collected were analyzed or interpreted because the task fell to a staff member in another department. As one practitioner explained, “I hand [our grants person] a stack of questionnaires each year, and they go through them and collate the comments.”

When asked how they reported their evaluation results, participants mentioned using a number of differing reporting methods that varied in formality. For example, all the Heads of Education I spoke with described being responsible for contributing to their institution’s annual report (see for example the education and wellness section of the 2016–2017 report produced by the Montreal Museum of Fine Arts, 2017a). Ten practitioners said that either they personally or in collaboration with other staff members prepared reports for those funders that required them demonstrating, among other things, that monies were spent according to their terms of evidence. While three practitioners described presenting evaluation findings via oral briefings during senior management meetings where no actual report was generated, one practitioner said that he prepared formal reports for his Board of Directors on a quarterly, annual, and ad-hoc basis. This included an evaluation report prepared by a qualified evaluator. Two practitioners further reported disseminating evaluation findings in peer-reviewed academic journals. Both did so in collaboration with external experts.

**Type of information collected.** Given the fact that most of the practitioners I spoke with said that their funders, not unlike their institutions, placed emphasis on output evaluation, it is perhaps unsurprising that roughly half reported collecting information on program activities such as the type and number of programs provided and the number of visitors served in addition to information on participant satisfaction and revenue generated. “Our evaluation is pretty basic so it’s mostly satisfaction and just did this work or not . . . that’s the extent of the evaluation we do
for programming,” said an internal researcher.

The other six practitioners interviewed reported collecting outcome or impact information in addition to information on program outputs. Two practitioners who worked at larger-budget institutions conducted outcome evaluations both independently and in collaboration with external experts. The remaining four practitioners either contracted out such evaluations or partnered with external experts in their undertaking. Some of the outcomes/impact evaluations mentioned included:

- a practitioner-led evaluation of the effectiveness of an interpretive space and an evaluator-led evaluation of the benefits of family programming,
- a practitioner-led evaluation of the impacts of outreach programs in addition to both a researcher-led evaluation of the impact of a wellness program on people with eating disorders and a researcher-led evaluation of the impact of community programs on vulnerable-sector persons’ feelings of belonging,
- a doctoral student-led evaluation of the effectiveness of a docent program,
- a researcher-led evaluation of the impact of a program targeted at Alzheimer’s patients, and
- a collaborative evaluation of the benefits of a wellness program on people living with dementia and their caregivers.

**Who was involved in the evaluation process.** The findings of the first phase of this study revealed a low level of visitor involvement in the evaluation process. In the vast majority of cases, program participants acted primarily as providers of data. In addition to serving as data sources, two practitioners reported that they further engaged visitors in the writing of interpretive text and sought their input in the front-end evaluation of both exhibition and educational
programs. “One of the things we do is just bring draft text panels down and have visitors read them out loud because they’re all so brilliant at writing these text panels we think,” said one Head of Education. Another one noted: “I would say that we do a lot of consulting with groups. I would say that we do a lot of consulting with artists and art historians around messaging and that kind of stuff. In terms of working with visitors, I think there is more that we could be doing.”

Despite the low level of visitor involvement in evaluation, Canadian art museum practitioners reported working with numerous other individuals on evaluation studies from internal staff across varying museum departments to a wide range of external experts. While in one case an internal researcher held primary responsibility for evaluation, the remaining practitioners I spoke with said that the task of evaluating programs fell to education staff (e.g., department heads, program managers, gallery educators, student interns). Roughly half of these practitioners also reported being assisted by staff members working in development, visitor services, marketing, communications, and finance.

Five of the practitioners who participated in this study further reported partnering with various experts external to their institutions on the evaluation of educational programs. They included individuals affiliated with universities, hospitals, health agencies, not-for-profit health organizations in addition to evaluation and media firms. For example, one education staff member collaborated with a post-doctoral researcher, two nurses, two occupational therapists, a medical doctor, and a speech language pathologist on the evaluation of her museum’s new wellness program. While in at least two cases practitioners relied on funds from their institutions or funders to finance such studies, in other instances external experts either volunteered their services or came with their own external funding. Referring to a university-based researcher she was collaborating with on the evaluation of a community program, one education staff member
noted, “She’s doing [the evaluation] because it’s a research interest of hers, so we’re not paying her to do it.” Eight of the 13 practitioners I interviewed further reported either hiring or partnering with external experts to conduct membership, marketing, or exhibitions research. Again, these domains of visitor studies inquiry will be examined in greater depth in Chapter 5.

**Satisfaction with the frequency and quality of program evaluation activities.** One education staff member expressed being satisfied with the number, quality, and frequency of the program evaluations her department conducted. She explained:

> We all know that there are tools out there that could help us get better at gathering feedback and doing more formal evaluation, but given how many of us there are and the quality of programming that we are putting out I think we are pretty satisfied with the way things are at the moment.

The remainder of the art museum practitioners I spoke with expressed either being dissatisfied with or interested in improving the quality and quantity of their program evaluations. “I don’t think we do [evaluation] nearly enough as an education department,” said one Head of Education. An internal researcher meanwhile noted:

> There’s a lot we could be learning about our programs. For instance, with school programs we don’t know how the program fulfilled teachers’ curriculum needs because we don’t find out. So, in my dreams I would have the ability to evaluate some of our key programs that have been running for a while and impact a lot of visitors.

**Evaluation Capacity Building**

The interview questions associated with this framework dimension aimed to capture the degree to which Canadian art museums were actively engaged in building their evaluative
capacities and to determine the specific activities they used to do so. In addition to providing empirical support for the use of the ECB *a priori* master code and the two sub-codes that fall under it (i.e., *direct activities* and *indirect activities*) on my start list of codes (Appendix F), the results of the first phase of this study point to the fact that ECB is still a relatively unusual occurrence within the sector. A mere three practitioners described having been engaged in formal, direct ECB activities at their respective places of work. At the first of these art museums, which had an annual operating budget of over 90 million dollars, the motivation for engaging in ECB was linked to an internal researcher’s ongoing professional development needs. The decision to partake in ECB at the second museum, which had an annual operating budget of just over 60 million dollars, was rooted in the vision of the museum’s Executive Director, who was described as “a real champion of audience evaluation.” At the third, much smaller art museum, the impetus for ECB came from its Head of Education, suggesting that organizational culture may have a greater impact on whether such museums are likely to engage in ECB than contextual characteristics such as size. It likewise reminds us that evaluation advocates, as Sanders (2001) rightly pointed out, can be found at any level of an organization and not just at the top.

Interestingly, the last two practitioners reported participating in multi-day face-to-face training workshops with Marianna Adams—principal of a leading visitor studies firm based in Dallas, Texas (Audience Focus Inc., 2019). A Head of Education described what he and his colleagues learned as a result of participating in one of these workshops, which, in the case of his museum, occurred within an actual participatory evaluation project:

It was sort of a crash course on evaluation and visitor studies . . . We just learned about basic evaluation, like what’s the difference between front-end evaluation and the other
kinds and between qualitative and quantitative research . . . We got some training around how to write surveys, select participants, interview people, develop an interview protocol, run focus groups, and do things like prototype activities. So initially it was just capacity building within our department—we had to learn all that stuff and at the same time we had the consultant do that kind of stuff for us and we would be around and helping. We would do some evaluations on the floor, and she would be there to observe and to coach and stuff like that. She also exposed us to how other museums were doing visitor research, and she introduced us to colleagues in the States.

Akin to engagement in an evaluator/researcher-led evaluation, involvement in direct ECB activities reportedly had significant long-term impacts on participants. Most notably, unlike their counterparts without any formal pre- or in-service training in evaluation, these practitioners possessed thorough understandings of both the three visitor studies domains of inquiry (i.e., market research, scholarly research, and evaluation) and the four phases of museum evaluation (i.e., front-end, formative, remedial, and summative) and used correct terminology when describing the evaluation activities occurring at their museums. As a result of participating in direct ECB activities, they likewise described applying their newfound evaluation knowledge and skills to their daily work—taking control of evaluating their departments’ programs and promoting the value and utility of evaluation amongst their colleagues.

Many of the remaining art museum practitioners I spoke with reported having attended conference presentations on various topics related to evaluation at annual meetings of the Canadian Art Gallery Educators Association, the National Art Education Association (NAEA), and the Visitor Studies Association. “When I do go to conferences like the NAEA or other conferences, I will go to sessions on evaluation because I am always curious to see what other
people are doing and how they do it,” remarked one practitioner. Several others also commented on how important networking with colleagues in other museums was to their own professional development around evaluation. Referring to an internal evaluator working at a Canadian museum of art, world culture, and natural history, another practitioner noted:

So, that person and I used to talk quite a bit because I have a qualitative museology background so I know all the theory and the museum side of things, and she has a mathematical background and knows the quantitative side of things so it was a nice pairing.

In general, the capacities of those art museum practitioners without formal pre-service training in evaluation who had nevertheless participated in direct ECB were significantly higher than their counterparts who had solely attended conferences or engaged in networking activities. While they remain unexplored in great depth, these findings shed preliminary light on the kinds of in-service training activities that hold promise for developing art museum practitioners’ hard skills in evaluation.

**Practice-Based Challenges and Limitations to Capacity**

One of the goals of the first phase of this study was to better understand both the challenges that art museum practitioners face when evaluating their educational programs and the factors that can either limit or enhance their EC. The results of this research supported the use of the *challenges a priori* master code on my start list of codes (Appendix F) while also revealing five related sub-codes (i.e., *lack of resources and supports, lack of internal demand for/commitment to evaluation, practice-based challenges, the marginalization of educational endeavours, and other challenges*). In other words, the results revealed that, for practitioners working in the Canadian art museum sector, engaging in the process of evaluation was not free
of challenges including certain difficulties already well articulated by researchers in the museum field (e.g., Allen et al., 2007; Hein, 1998; Hooper-Greenhill, 2004; Michalchik & Gallagher, 2010; National Research Council, 2009).

For example, several practitioners described struggling with participant recruitment, particularly during 19+ networking events. As one Head of Education noted, “It’s been tough because there’s always loud music and it’s very much a party environment so it’s really hard to get people to participate in one-on-one interviews.” Others meanwhile commented on the difficulty of obtaining feedback from certain participant groups:

I have questionnaires that I send out, but I very rarely get comments from the participants, which is not surprising because many of them have mental health issues or addiction problems. And, because the bulk of my programs engage vulnerable populations, it’s a major issue for me.

Selecting which measures to use to evaluate the success of programs posed an additional challenge:

For us, like any institution, tracking attendance and revenue is really important, but there are times when we know a program might not draw as large a crowd, but how do you assess if one is more or less successful than another if you’re just measuring success based on attendance?

While some of the practice-based challenges reported here are to be expected when measuring learning in informal settings (e.g., inserting data collection into programs in a way that does not interrupt participants’ experiences), others (e.g., identifying outcome measures, formulating interview questions that elucidate quality responses) were indicative of art museum practitioners’ limited evaluative knowledge and skills. Despite this limitation, from the
perspectives of the practitioners I spoke with, a lack of evaluation resources (i.e., time, staff, money) was believed to pose the biggest challenge to their capacity for evaluation. Specifically, the majority reported being challenged by a lack of time followed by limited financial and human resources. “Evaluation is something that is not exactly a priority in a lot of cases just because there is no time to do it,” said one Head of Education. An internal researcher further noted, “There’s no shortage of desire for program evaluation, it’s just that we haven’t quite caught up in terms of how we resource it. Right now, we’re doing a lot of magic tricks.” This finding reinforces a recent observation made by Adams (2012) that, “while the ‘want to’ in the practitioner’s desire to conduct evaluation is much stronger than ever before, the ‘can do’ is a different matter” (p. 28).

The second challenge that practitioners mentioned most frequently as limiting their capacity for evaluation included a lack of internal demand for and commitment to evaluation. One Head of Education spoke to the former challenge when he said:

We don’t put a lot of effort into program evaluation, but I know why we don’t—it’s not a requirement, and therefore we don’t have time to do it. So, unless someone gives me capacity and prioritizes it, it’s not happening.

Together, these findings both serve to illuminate Canadian art museum practitioners’ evaluation struggles and illustrate the ways in which factors such as culture, demand for evaluation, and resources and supports can limit evaluation capacity in these professional settings.

Practitioners’ Self-Reported Evaluation Needs

Another purpose of this phase of the study was to determine what type of assistance art museum practitioners thought would enable them to increase their EC and more effectively evaluate their educational programs. The results of the findings of the first phase of this study
supported the use of the evaluation needs a priori master code on my start list of codes (Appendix F) while also revealing four other needs-based sub-codes (i.e., resources and supports, internal demand for and organizational commitment to mainstreaming evaluation, external demand for evaluation, and training in evaluation). Given practitioners’ self-reported challenges, it is unsurprising that the top need they identified was the need for additional resources and supports (e.g., human and financial resources, time, evaluation tools, technology).

Specifically, the overwhelming majority of practitioners said they wanted help in the form of human resources followed by, and often coupled with, increased funding. For example, some heads of education expressed feeling that being able to hire additional education staff would help free them up from their primary responsibilities sufficiently enough to engage in more evaluation. These practitioners nevertheless described needing access to evaluation reports from other museums in addition to examples of evaluation techniques and instruments to guide them in this kind of work.

Other practitioners less willing to go it alone perceived benefitting from either the assistance of external experts (e.g., professional evaluator, university researcher) or in-house research staff. Such assistance was viewed as having the potential to increase an art museum’s overall capacity to conduct evaluations as well as the credibility and influence of their findings. “I think we could do more if we had a dedicated internal position who could oversee evaluation, not just for education, but across the board,” said one Head of Education. Another practitioner noted, “If it was to be done by a qualified and trained professional evaluator then maybe the results would be different and evaluation would have a lot more impact.”

An additional frequently identified practitioner need was the need for increased institutional demand for and support of evaluation. “If there was a requirement as part of our
quarterly reporting, for example, to the Board then evaluation would happen,” noted one Head of Education. Not only did practitioners wish to see leadership request and use evaluation data routinely in their governance, but, more importantly, they wanted to see evaluation integrated into their museums’ everyday operations in sustainable ways—a process Sanders (2002) coined *mainstreaming evaluation*. For example, one Head of Education expressed her desire for a whole organization approach to evaluation by saying, “The prioritization of evaluation work needs to be folded into every facet of the gallery. It needs to be a gallery-wide commitment.” Another practitioner described thinking that evaluation should be mainstreamed not only vertically (i.e., at all levels of the organization), but also horizontally (i.e., across departments through appropriate structures) (Picciotto, 2002):

> The drive for evaluation should be really be coming from all levels in the institution from the Board, to the Director and management teams, to individual staff, and it should be a part of the strategic plan, a part of what we do in all departments.

Reported much less frequently, a number of practitioners also felt that Canadian funders could play a greater role in supporting evaluation in art museums by requiring them to direct a portion of a total grant’s value towards the comprehensive evaluation of programmatic impacts. “A requirement from funders would be the best way to ensure that evaluation happens,” argued one Head of Education.

Although a lack of evaluative knowledge and skills was not necessarily seen by practitioners as posing a challenge to their capacity for evaluation, the art museum practitioners I spoke with all agreed that they would benefit from training in evaluation. As one Head of Education noted, “Some training would definitely be very empowering and not just for me, but for all of my staff.” Specifically, practitioners expressed wanting to (a) increase their knowledge
of differing social science research methods and evaluation approaches and (b) gain skills related to both designing instruments (e.g., questionnaires and interview protocols) and collecting, analyzing, and interpreting qualitative and quantitative data alike. Some further expressed wanting to leverage social media and mobile applications for data collection while others wished to learn how to use online survey tools such as Survey Monkey or various spreadsheet software (e.g., Microsoft Excel) to calculate descriptive statistics for their datasets. Those practitioners with backgrounds in visitor research and evaluation likewise saw value in continued professional development. One internal researcher, for instance, was in the process of pursuing additional certification during the first stage of this research.

Practitioners’ evaluation resource needs combined with their desires for evaluation mainstreaming suggest that ECB in the context of art museums should be directed at the institutional level with the ultimate goal of integrating evaluation as a core organizational value. Additionally, because regardless of whether an art museum is committed to evaluation or not, it is often education staff who implement it, ECB efforts in these settings need to simultaneously be directed at the individual level to meet practitioners’ self-reported evaluation training needs.

**Reactivity and Opportunities for the Future**

Emerging from my analysis of the Phase One interview data was an additional master code not anticipated at the beginning of the study—*reactivity and opportunities for the future*—which was further divided into two sub-codes (i.e., *reactivity* and *opportunities*). Specifically, the first phase of this study resulted in a key instrumentation effect among several of the art museum practitioners I spoke with. Namely, for these practitioners, the interview process stimulated in them a renewed commitment to evaluation. “I think what I’m going to try to do is find some money for next year to hire an evaluator,” said one Head of Education. Another noted:
When we built the (learning centre), we put a research fellow into the budget, and we have been unable to find funding for that. So, this conversation is reminding me that I need to dig up that proposal and start flying that flag again because that would allow us to do that program evaluation which we need to get done.

In addition, numerous practitioners talked about what they saw as opportunities for enhancing their art museums’ future evaluation practices and capacities. One such opportunity included forging museum-university partnerships around visitor studies and evaluation. “More partnerships with universities would really be the most beneficial I think,” said one museum staff member. Owen and Visscher (2015) agree that there is great value in this approach. Speaking about collaborating with graduate students specifically, they wrote, “Conducting audience research through university and museum partnerships has the potential to be mutually beneficial—providing museums with valuable insights into the audiences they serve and giving students hands-on experience in a professional setting” (p. 71). Indeed, as the Phase Two findings illustrate, for those art museums interested in integrating evaluation into their ongoing work, museum-university partnerships can be a viable strategy (see Case Portrait 2: The Montreal Museum of Fine Arts).

Practitioners identified the building of museum networks as another potential opportunity for enhancing their EC. Pointing to the collaboration of twelve museums in Balboa Park, San Diego as an example, one Head of Education commented on how this strategy can offer art museums a distinct starting point for overcoming limitations of resource scarcity:

There’s a park in San Diego with a number of small museums who kind of group their resources together to do audience evaluation—Balboa Park I think it is. They probably
don’t have the resources to conduct evaluations independently, but if a number of them get together, they can.

Another practitioner further noted the strategy’s potential for carrying out pan-institutional evaluations:

What if four or five museums got together and decided what are two or three big evaluation questions that they would love to have answered and then they worked independently to explore those questions and later compared their data sets? Like the Vancouver Art Gallery, the Montreal Museum of Fine Arts, the National Gallery of Canada, and the Art Gallery of Ontario because they’re all of similar size. Or, just the Ontario art museums banding together to try to better understand their region. I think there’s real power in collaboration, and there’s so much you can learn from your own data set in comparison to someone else’s!

Steele-Inama (2015) believes that forming evaluation networks for museum professionals can “positively influence evaluative thinking, implementation, and use” (p. 80). Museum networks the likes of Balboa Park and the Denver Evaluation Network (DEN, 2019) can therefore serve as models for those Canadian art museums looking to increase their capacities for evaluation.

**Chapter Summary**

The findings of the first phase of this study indicate that, precluding a few minor adjustments (see Chapter 6), EC in the Canadian art museum sector can in fact be described through the seven dimensions included in the *Conceptual Framework of Evaluation Capacity in Art Museums*. For instance, considerable evidence was found to suggest that factors such as *context, culture, demand for evaluation*, and *resources and supports* can contribute to either
promoting (e.g., when leadership champions evaluation) or limiting an art museum’s overall capacity for evaluation. It is important to note that, while presented separately, the dimensions and sub-dimensions in the framework are highly interdependent in practice as illustrated by the observed links between an art museum’s evaluation culture, learning climate, and its organizational structure.

In addition to providing validity evidence for those dimensions conceptualized as influencing factors in my conceptual framework, the findings from this study likewise provide significant insight into Canadian art museum’s evaluative practices, their current capacities to evaluate their educational and public programs, and their level of engagement with ECB. Overall, the interview results painted a portrait of moderate capacity for evaluation across the sector with smaller pockets of high capacity. EC was found to be most developed at the Montreal Museum of Fine Arts and the Art Gallery of Ontario—museums with annual budgets of 40 and 90+ million respectively. Both art museums were selected for case study. Moreover, the findings of this study highlight the challenges Canadian museums face on their journeys towards evaluating programmatic impact and their needs for assistance around evaluation. Practitioners’ self-reported challenges and needs suggest that the availability of evaluation training, resources, and other supports for those tasked with program evaluation is of vital importance in art museums.
Chapter 5

Phase Two Findings

This chapter reports findings from two case studies of Canadian art museums that, based on the Phase One findings, were identified as operating at the highest level of capacity for evaluation in the country—the Montreal Museum of Fine Arts and the Art Gallery of Ontario. The case studies focused on these museums’ overall capacities to conduct a wide range of visitor studies (i.e., scholarly research, market research, and both exhibition and educational program evaluation) and use the findings of those studies for varied purposes. Embedded in this focus was an interest in understanding how these museums’ program evaluation capacities compared to their capacities to engage in other domains of visitor studies inquiry. A case portrait (Lightfoot, 1983) of each participating museum is presented first. Akin to the presentation of the Phase One results, findings for each individual art museum are discussed according to a mix of both a priori and, where applicable, emergent themes. The individual case portraits are followed by the presentation of cross-case results which are discussed according to emergent themes.

Case Portrait 1: The Montreal Museum of Fine Arts

Introduction and study context. Occupying a total floor area of 53,095 square metres, the Montreal Museum of Fine Arts (MMFA) is Canada’s largest art museum and the 18th largest in North America (MMFA, 2017a). The MMFA’s primary mission is “the enhancement, conservation, and presentation of its encyclopedic collection, the presentation of major exhibitions, and the creation of various educational, art therapy, and social/community programs” (MMFA, 2019a, para 1). Further to this, the institution positions itself as a caring, humanist, and socially-engaged museum—one that sees itself as playing a vital role in responding to broad societal issues (e.g., mental disabilities, racism, homophobia, illiteracy,
school dropout rates, poverty, homelessness, bullying, violence, radicalization, alienation, aging, etc.) and helping to drive social change. The MMFA’s humanist and inclusive values are particularly reflected in both the museum’s unique programming and research initiatives discussed in greater detail in the next section.

The MMFA’s collection includes more than 41,000 works spanning from antiquity to the present day that are displayed across five separate pavilions. These pavilions include the Michal and Renata Hornstein Pavilion devoted to archaeology and ancient art, the Lilian and David M. Stewart Pavilion dedicated to decorative art and design, the Jean-Noël Desmarais Pavilion featuring modern and contemporary art, the Claire and Marc Bourgie Pavilion, which both houses the Quebec and Canadian art collection and includes a 444-seat concert hall, and the new Michal and Renata Hornstein Pavilion for Peace—home to the museum’s collection of international art.

In the last fiscal year, the museum reported welcoming a record-breaking 1.3 million visitors, saw roughly 301,170 of them take part in its educational and community programs, and boasted a VIP base numbering over 109,010 members, making it the top-ranked art museum in Canada for both attendance and membership (MMFA, 2017a). As a not-for-profit organization, the MMFA relies on a mix of self-generated revenues and financial support from the public. Specifically, 54% of its annual operating budget comes from such sources as ticket and membership sales in addition to donations and sponsorships while its remaining revenue comes from government grants (e.g., Le Ministère de la Culture et des Communication, the Department of Canadian Heritage, the Canada Council for the Arts, the Conseil des arts de Montréal, and the City of Montreal) (MMFA, 2017a, 2017b).
Culture. The MMFA has a high capacity for organizational learning with an emphasis on promoting collaboration and innovation. It was noted by interviewees as being, among other things, “bold,” “flexible,” “responsive,” “open-minded,” “ingenious,” “progressive,” “multi-disciplinary,” “creative,” and “dynamic.” It was further described by Nathalie Bondil, its Director General, as a “testing ground for innovation” that was “continually seeking to reinvent itself” and known for both “thinking outside of the box” and “a thirst for experimentation without fear of failure.” This culture of learning permeates all the MMFA’s endeavours and is perhaps most notable in the areas of art museum education and visitor studies, which are both integral to the museum’s mission and overall commitment to social inclusion and change.

Art museum education. The 2016 opening of the Michel de la Chenelière International Atelier of Education and Art Therapy—the largest educational facility of any North American art museum (MMFA, 2017a)—is a testament to the MMFA’s commitment to art museum education. Occupying two of six levels of the Michal and Renata Hornstein Pavilion for Peace, the atelier includes two multipurpose areas, a 300-seat lunch room, 12 studios, two venues for multidisciplinary activities (i.e., the Bourgie Concert Hall and Maxwell Cummings Auditorium), a dedicated area for groups taking part in clinical research projects, a consultation area for doctors, and an open art therapy studio. These facilities have enabled the MMFA’s Education and Wellness department to reach a greater number of K–12 students, families with children, adults, and seniors through both its educational and public programs than ever before. They have likewise enabled the department to broaden the scope of its art museum education activities to include another two key programming areas (a) accessibility and inclusion and (b) art therapy and well-being. Each of these three programming areas is discussed in greater detail below.
The Education and Wellness department offers an expansive schedule of educational and public programs aimed at a wide range of audiences including traditional school tours and workshops (MMFA, 2019b) and a separate website designed for use as a pedagogical resource by teachers (MMFA, 2019c). The department’s public programs range from creative workshops for organizations (MMFA, 2019d), to yoga sessions for seniors (MMFA, 2019e), to drop-in weekend programming for families and day camps for youth (MMFA, 2019f). It likewise includes activities aimed at adult audiences such as lectures, film screenings, courses and workshops, guided tours, musical events, and conferences (MMFA, 2019g).

The Education and Wellness department’s accessibility and inclusion programs were designed to enable those visitors who have typically been excluded from cultural experiences (e.g., vulnerable persons, immigrant women, young single mothers, youth at risk, vulnerable and illiterate persons, etc.) to take part in free activities specifically tailored to them. Of note are two programs. The first, offered in collaboration with the Accueil Bonneau\textsuperscript{15}, hopes to foster a sense of belonging among men suffering from various problems related to homelessness (MMFA, 2019h). The second program, a joint effort of the MMFA and Leave out Violence Quebec, seeks to reduce the effects of verbal, physical, or sexual abuse on the lives of adolescents (MMFA, 2019h).

Lastly, the Education and Wellness department’s art therapy and health and well-being programs were designed to promote the physical and mental well-being of a wide range of groups from persons living with eating and speech disorders, arrhythmia, or intellectual disabilities to those who have been affected by suicide or are struggling with mental health

\textsuperscript{15} The Accueil Bonneau is a nonprofit organization that provides shelter and other services to those who find themselves homeless in Montreal.
issues. Operating in collaboration with the Douglas Mental Health University Institute\textsuperscript{16}, \textit{Sharing the Douglas} is an example of an art therapy program for persons living with anorexia and bulimia that aims to shift participants’ perceptions of their body image (MMFA, 2019i, 2019j). Examples of the museum’s programming in health and well-being include two programs offered in partnership with each the Montreal Heart Institute and the Sainte-Justine University Hospital Center.\textsuperscript{17} The goals of the former program are to improve the physical health of patients living with high heart rates while the latter was designed to promote wellness among young people suffering from some form of psychopathology (MMFA, 2019j, 2019k). Although few art museums today would be content with a traditionally narrow definition of art museum education—one focused on school programming alone—the MMFA’s art therapy and health and well-being programs provide an expanded vision for the type of programming that can occur in these unique learning settings.

\textit{Visitor studies}. The MMFA’s commitment to art museum education was paralleled in its dedication to the three inquiry domains of museum visitor studies (i.e., market research, scholarly research, and evaluation), with special focus placed on both applied research and program evaluation. Nathalie Bondil noted that over the last 18 years the MMFA has become “a real research laboratory” designed to systematically measure the impacts of its programs, especially its art therapy and health and well-being programs, on participants. “I’m convinced that art is good for you. And we intend to prove it scientifically,” she said. Although there was no dedicated evaluation position at the museum, the MMFA consistently partnered with a wide range of external experts (e.g., researchers, graduate students, physicians, therapists, etc.)

\textsuperscript{16} The Douglas Mental Health University Institute is a psychiatric hospital in Montreal affiliated with McGill University.

\textsuperscript{17} Affiliated with the University of Montreal, the Sainte-Justine University Hospital Center is the largest mother and child university health centre in Canada.
working in local universities, health institutes, and nonprofit organizations to examine the benefits that can be attributed to both art and art museum education (broadly defined). “It’s a new way of doing research, and it’s a new way of collaborating with different people,” explained Thomas Bastien, Director of Education and Wellness.

At the time that this research was conducted, the MMFA was involved in no fewer than 32 collaborative research/evaluation studies—five of which were focused on its accessibility and inclusion programs, 13 of which were focused on its educational and public programs, and a further 14 that focused on the museum’s art therapy and health and well-being programs. The MMFA’s research into these programs was supported by the work of numerous transdisciplinary advisory committees that were comprised of various experts who are leaders in their respective fields. These committees met several times a year to observe studies that were underway and provide the museum with direction on its priorities for research and evaluation.

**Demand for visitor studies.** At the MMFA, visitor studies were driven less by external requirements and more by their integration into the museum’s learning culture. While interviewees identified different reasons for undertaking market research, scholarly research, and evaluation depending on their role and the department they worked for, seven key interrelated drivers of visitor studies emerged from their comments. First, as a strong advocate of scholarly research, interviewees noted the museum’s Director General as being a principal driving force behind expectations for this particular domain of visitor studies inquiry at the MMFA. Another driver was related to the museum’s fundamental values. Namely, interviewees reported that the museum’s desire to be “at the service of society” (MMFA, 2016, p. 22) by helping to promote social cohesion and individual health and well-being was a significant motivator for its applied
research and program evaluation activities. “Carrying out research is at the core of the humanist museum’s mission,” explained Nathalie Bondil.

Interviewees reported the museum’s deep culture of innovation as another core reason behind the demand for visitor studies information at the MMFA. Specifically, its quest to spur “novel ideas” (MMFA, 2016, p. 26) and foster “inventive solutions” (MMFA, 2016, p. 26) to a wide range of societal ills has led the museum to carry out “pioneering research” (MMFA, 2017a, p. 26) through “innovative forms of cooperation” (MMFA, 2017a, p. 8). The fourth driver of visitor studies at the MMFA was reported as being the information needs of senior managers and program officers. Referring to market research activities at the MMFA, Jean-Sébastien Bélanger, its Head of Membership and Customer Services, noted, “Conversion marketing is the membership department’s core business, and to convert museum visitors to members and members to donors, you need data. It’s the most important thing for me to be successful in my job.” These same data were likewise described as being critical to the work of the museum’s communications department. Pascale Chassé, the department’s Director explained, “You have to know your visitors if you want to reach them.”

As evidenced by comments made by Stephen Legari, Educational Programmes Officer-Art Therapy, a desire for continuous improvement—particularly in the area of program development—was yet another driving force behind visitor studies at the MMFA. He stated that evaluation information enables him to “like a mechanic, fine tune programs over and over again” and that this type of program monitoring and tailoring has been “invaluable in the evolution of the museum’s programs.” The last internal driver of visitor studies at the MMFA was its desire to contribute to knowledge in fields such as museum education, health, and the creative arts therapies and share examples of good practice with other art museums both in Canada and
abroad. Thomas Bastien described the museum’s commitment to both contributing and mobilizing knowledge by saying:

What we are doing at the museum can be reproduced in other museums, and that’s the reason that we have research going on—because we want to create protocols that can be replicated in other museums around the world. That’s the reason that the team travels so much to do presentations in France, in the United States, in Spain and in other countries in Europe. It’s because we have been creating expertise that can be shared with other institutions, and that’s why research is really relevant in the museum.

While visitor studies at the MMFA were primarily motivated by internal forces such as those described above, this is not to say that accountability did not also play a role. Many of the museum’s funders and donors have a keen interest in knowing how their monies are being used. Referring to meeting funders’ and donors’ external demands for evaluation, Thomas Bastien noted, “It’s about credibility. It’s about letting them know that we used the money in the right way.”

The type of information interviewees were required to collect differed based on the domain of visitor studies inquiry that they engaged in. For example, those museum practitioners who held primary responsibility for applied research and program evaluation at the MMFA were largely expected to identify and measure the outcomes of programs such as the impact of art on well-being. Practitioners who were responsible for market research and exhibition evaluation, on the other hand, generally collected output information (e.g., demographic information, participant satisfaction, amount of revenue generated through ticket sales, etc.), illustrating that different types of information are tied to differing functions of an art museum.
Resources and supports. The MMFA did not have a dedicated evaluation unit or position. Instead, visitor studies were undertaken by staff working in various departments who relied on varying degrees of assistance from outside experts depending on the domain of inquiry in question. The MMFA’s Education and Wellness department was responsible for initiating a wide range of applied research and program evaluation studies in collaboration with recognized experts working in the fields of education, art therapy, health, and the social services. For instance, in one case (Thaler et al., 2017), an educational program officer undertook an evaluation of one of the museum’s art therapy programs in partnership with the following outside experts: Dr. Howard Steiger (Professor, Department of Psychology, McGill University/Chief, Eating Disorders Program, Douglas Mental Health University Institute), Dr. Lea Thaler (Assistant Professor, Department of Psychiatry, McGill University/Psychologist, Eating Disorders Program, Douglas Mental Health University Institute), Dr. Josée Leclerc (Associate Professor, Department of Creative Arts Therapies, Concordia University), Nadine Ferenczy (Psychoeducator, Eating Disorders Program, Douglas Mental Health University Institute), Catherine-Emmanuelle Drapeau (Doctoral Candidate, Psychology Department, Université du Québec à Montréal), and both Danaelle Cottier and Esther Kahan (Research Associates, Montreal West Island Integrated University Health & Social Service Centre).

This partnership approach to research and evaluation has allowed the department to overcome two important limitations of resource scarcity that, as the Phase One results have indicated, pervade the Canadian art museum sector (i.e., a lack of staff time and sufficient funds for evaluation). It likewise enabled the Education and Wellness department to carry out more studies than it could have otherwise conducted in an internal capacity while simultaneously diversifying its research expertise. Thomas Bastien explained:
And when you do decide to do research, you need to first conduct a review of the existing literature and that takes a lot of time, and we don’t have that kind of time here. Plus, if I wanted to do the kind of research that we are currently doing at the museum internally, I would have to hire a number of different researchers, and I don’t have the capacity to do that.

As mentioned earlier, the MMFA’s applied research and evaluation activities were supported by the work of several voluntary advisory committees (i.e., the Arts and Education Advisory Committee, the Autism and Education Advisory Committee, the Diversity and Inclusion Advisory Committee, and the Art and Health Advisory Committee). The MMFA’s new Art and Health Advisory Committee, for example, was comprised of 17 individuals representing the MMFA in addition to numerous local and international universities, foundations, granting agencies, and nonprofits (MMFA, 2017c, 2019i, 2019j). Among its members were Rémi Quirion (Quebec’s Chief Scientist), Isabelle Peretz and Robert J. Zatorre (Co-directors, the International Laboratory for Brain, Music and Sound Research, Université de Montréal), Olivier Beauchet (Dr. Joseph Kaufmann Professor of Geriatric Medicine, McGill University), Josée Leclerc (Associate Professor, Department of Creative Arts Therapies, Concordia University), Marie Gaille (Deputy Scientific Director, Institute of Human and Social Sciences, Université Paris-Diderot), Philippe Walker (Chief Scientific Officer and Founder, The NEOMED Institute18), and Her Excellency the Right Honourable Julie Payette (Governor General of Canada). Together, these committee members shared their expertise on best practices in their respective fields, evaluated the merits of potential research projects, and provided a vision for the MMFA’s

18 The NEOMED Institute is a nonprofit, public-private institution based in Quebec that aims to bridge the gap between basic research and the commercialization of new drugs.
research and evaluation agenda as well as suggestions for potential partnerships.

In addition to receiving significant research and evaluation support from external experts and volunteer advisory committee members alike, the Education and Wellness department was also assisted in such tasks by various staff members working across the museum. For example, the Membership and Customer Services department assisted Education and Wellness staff in the evaluation of educational spaces that were created for three separate major exhibitions. Jean-Sébastien Bélanger explained:

We’ve been doing this for a few exhibitions now where we’ve included little sections for kids. [The Head of Education and Wellness] wanted to know how many people were using those sections so in my exhibition exit surveys I will be asking visitors that question, and I will be giving him that data so he’ll know approximately how many times someone went into that section, and he’ll be able to use that number to report back to the donor that gave the money for that section to exist.

While the Education and Wellness department was responsible for leading the MMFA’s applied research and program evaluation agenda, the Communications and Membership and Customer Services departments were responsible for conducting both market research and exhibition evaluation at the museum. The heads of both departments assumed primary responsibility for these domains of visitor studies inquiry, but also enlisted the help of external experts in their work albeit to a lesser degree than the Education and Wellness department. “I work with a media firm that has the tools to analyze postal codes, and they analyze for me where our visitors are coming from,” said Pascale Chassé by way of example. She likewise described working with graduate students at the Université du Montréal on a bi-annual large-scale visitor survey. Across the foregoing three museum departments, staff reported using the MMFA’s
computerized database to collect, analyze, and manage information. Jean-Sébastien Bélanger noted:

We’re work with a data base that is shared by most departments, including education, communications, public relations, membership, the foundation, and rentals. So, each department collects different data, but the data that are collected are all in that data base.

**Capacity to do visitor studies.** At the MMFA, not unlike at other Canadian art museums, the interviewees I spoke to during the second phase of this study reported having come to their respective positions from a wide range of academic disciplines. These included art history, French literature, and combinations of engineering, genetics, the fine arts, educational psychology, political science, human resources, and marketing and communications. Of the five museum practitioners I spoke with, one had received formal training in survey research as part of his professional preparation while another had specialized in program evaluation during her graduate studies. Others still had experience with both basic and clinical research.

Those interviewees with formal training regularly drew on their knowledge and skills in areas such as participant observation and basic survey design and analysis in their work. This said, they were also quick to recognize their own limitations and frequently relied on partnerships with outside experts (e.g., graduate students, university professors, health professionals, researchers, consultants) to bolster their own internal capacities for visitor studies. For example, the Education and Wellness department’s significant capacity for evaluation and applied research was owed less to the knowledge and skills of individual staff members and more to the museum’s broad-based partnership approach. Thomas Bastien explained how the strategy of partnering allowed the MMFA to capitalize on diverse expertise in both program development and research and evaluation:
We collaborate with researchers, but we also collaborate with various associations in Montreal, which means that we have expertise in museum work, our collaborating associations have experience working with the clientele they serve, and the researchers have research expertise. So, when you combine those three things together, it’s amazing what we see happen.

This finding suggests that art museums do not necessarily need to possess high levels of technical expertise internally to undertake systematic data-based inquiries into their programs. Rather, interviewees thought that relying on such expertise from outside their own organizations was a powerful way for them to enhance their evaluative efforts (see the Visitor Studies Activity section for a concrete example).

**Capacity to use visitor studies.** The MMFA used the findings of visitor studies in a number of different ways depending on the department in question. The MMFA’s Foundation reported using the information that was generated by the Membership and Customer Services department to make a case for new and continued financial support to prospective or existing sponsors and donors. The Communications department primarily used the results of the market research it conducted to identify target audiences and develop, monitor, and evaluate advertising campaigns aimed at those audiences. The MMFA’s Membership and Customer Services department, on the other hand, used the customer data it collected to both acquire new members and donors and manage their relationship with the museum through targeted communications strategies. Jean-Sébastien Bélanger explained:

If you are a member and you have not been to the museum for six months, we will try to push you to come to the museum because we want you to renew your membership, and you won’t renew if you don’t come. So, [this information] helps us drive your visit to the
museum by sending you an email or even giving you a call. We also look at what emails members open. If they don’t open emails, they’ll receive a postal card instead reminding them that their membership is lapsing. If they do open emails, we’ll send them an email reminders, and if we see that they opened it, but did not renew, we’ll look at which emails work and which don’t work.

Visitor research and evaluation findings were also reported as being used by the department to detect instances of fraud, improve exhibition programs, and report to the museum’s governance.

The MMFA’s Education and Wellness department reported using the results of the applied research and program evaluation studies it conducted first and foremost to demonstrate programmatic impact and contribute to the generation of new knowledge in fields such as education, art therapy, and the health sciences. Members of the department likewise reported using the findings of such studies to validate the museum’s programming models. “It makes me feel more confident that our program protocols—these different ways of delivering art therapy at the museum—are working,” Stephen Legari explained.

Further to these examples of the conceptual use of findings, evaluation and research knowledge was also reported as being used for a number of other parallel purposes from program development, monitoring, and improvement to accountability. Specifically, interviewees described using findings to make staffing decisions; adjust the pacing, duration, and order of educational activities; determine which pilot projects should be rolled out as ongoing partnerships; promote programs in the media; and both demonstrate program relevance to funders and make the case for their continued support.

While education staff members did not report learning to think more empirically or acquiring practical evaluation skills as a result of participating in expert-led evaluation and
research studies—perhaps on account of the fact that these processes were not used to intentionally build capacity or the fact that the MMFA already has an established evaluation culture—they did note two other process use outcomes. These included using evaluation and research processes to (a) enhance shared understandings of programs among those involved and (b) both reinforce and strengthen program impacts (Patton, 2008).

Visitor studies activities. As has already been well established, the MMFA engaged in a wide range of visitor studies activity. The Communications and Membership and Visitor Services departments conducted market research and exhibition evaluation studies to understand audiences’ perceptions, attitudes, demographics, and visiting patterns as well as to determine the overall effectiveness of the MMFA’s advertising campaigns, communications plans, and exhibition programs. The Education and Wellness department, on the other hand, primarily conducted applied research and program evaluation studies focused on examining the impacts of the museum’s educational, health and wellness, and inclusivity and accessibility programs.

Market research and exhibition evaluation. Those interviewees with responsibility for conducting market research and exhibition evaluation at the MMFA reported being engaged in a process of continuous data collection for monitoring, tailoring, and conversion purposes (i.e., to convert visitors to members and members to donors) while also conducting periodic large-scale surveys for accountability purposes with the assistance of either consultants or graduate students. Overall, the information they reported collecting was mostly output-based in nature. Specifically, interviewees with responsibility for these domains of visitor studies inquiry collected sociodemographic data (e.g., education level, location, income, etc.) as well as data on both the nature and frequency of clients’ visits and their levels of satisfaction with various museum experiences. Such data were most frequently gathered using document reviews (e.g., attendance
and ticketing records) and questionnaires while Microsoft Excel was most often utilized to calculate descriptive statistics for quantitative data sets. The findings of market research and exhibition evaluation studies were both disseminated in the museum’s annual report and made available to the museum’s Foundation to share with those donors and funders that required them.

**Applied research and program evaluation.** Interviewees from the Education and Wellness department reported carrying out informal, improvement-oriented, mostly output-based formative evaluations of their programs themselves as well as partnering with one or more external experts to conduct more systematic, judgment-oriented, outcomes-based summative studies of their programs. Table 6 provides an overview of a selection of the MMFA’s ongoing applied research and evaluation projects. Included in this overview are the names of the institutions, organizations, and/or associations that the museum partnered with on each study.

Table 6

**Overview of select ongoing research and evaluation projects**

<table>
<thead>
<tr>
<th>Focus of Study</th>
<th>External Partner(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact of museum visits and workshops on people with eating disorders</td>
<td>Concordia University, McGill University, Douglas Mental Health University Institute,</td>
</tr>
<tr>
<td>Impact of museum visits and workshops on youth aged 14–25 suffering from psychiatric disorders</td>
<td>Sainte-Justine University Hospital Center</td>
</tr>
<tr>
<td>Impact of museum visits on patients with cardiac arrhythmia</td>
<td>Montreal Heart Institute</td>
</tr>
<tr>
<td>Impact of museum visits on people living with Alzheimer’s</td>
<td>McGill University, Alzheimer Society of Montreal</td>
</tr>
<tr>
<td>(Neuro)physiological impact of museum visits and workshops on people on the Autism spectrum</td>
<td>Université de Montréal, Laboratoire de Neurosciences Cognitives at the Centre National de la Recherche Scientifique (Paris, France)</td>
</tr>
<tr>
<td>Impact of museum activities on people aged 65 and older</td>
<td>RUIS McGill Centre of Excellence on Longevity, Jewish General Hospital</td>
</tr>
<tr>
<td>Impact of museum art therapy programming on breast cancer patients and survivors</td>
<td>Quebec Breast Cancer Foundation, Université du Québec en Abitibi-Témiscamingue</td>
</tr>
<tr>
<td>History and evolution of health and wellness programs at the MMFA</td>
<td>Université du Québec à Montréal</td>
</tr>
<tr>
<td>Impact of museum visits and workshops on epilepsy patients</td>
<td>Concordia University, McGill University</td>
</tr>
</tbody>
</table>
Table 6 Continued

<table>
<thead>
<tr>
<th>Focus of Study</th>
<th>External Partner(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact of museum art therapy programming on voluntary immigrants</td>
<td>Université du Québec en Abitibi-Témiscamingue</td>
</tr>
<tr>
<td>Effectiveness of a new docent training program</td>
<td>Concordia University</td>
</tr>
</tbody>
</table>

When conducting formative studies, interviewees most often mentioned using data collection methods such as document reviews (e.g., of attendance records), questionnaires, direct observations, and informal conversations. The information collected via these means was typically analyzed using either qualitative content analysis or basic quantitative assessment techniques. While reported less frequently, some interviewees also described relying on artifact reviews—particularly of artwork created by participants engaged in the museum’s art therapy and ÉducArt programming. “From an art therapy perspective, one of the primary ways that we notice change in participants is through the language of art making,” Stephen Legari noted. He went on to explain:

If somebody comes in as a participant for the 20th time and their art starts to take on a different shape and form and level of experimentation, that to me is information about some degree of comfort and some degree of flexibility of expression. When people start to leave their artwork at the museum because they want to come back to it the following week, that gives me an indication that they feel a sense of belonging at the museum. That’s it’s become a homed space. If, on the other hand, somebody sticks very rigidly with a single material or a single approach and/or technique, while it may be very pleasurable and calming to them, it may also show that they are not able or willing to transcend.
Across the systematic data-based inquiries conducted in partnership with external experts, the data collection, analysis, and interpretation methods that were employed varied based on the nature of the study in question. While these methods are too numerous to recount here, I point to Thaler et al.’s (2017) study by way of example because it is illustrative of the MMFA’s partnership approach to both conducting research and evaluation studies and disseminating their results. As Thomas Bastien pointed out, “90% of the time [MMFA staff] are co-writers of the research that has been done at the museum.”

In Thaler et al.’s (2017) study, the authors—two of whom worked at the MMFA—examined both the suitability and short-term impacts of the museum’s art therapy programming on 78 adult patients from the Douglas Institute’s Eating Disorders Program. At the end of each of their 11 visits to the museum, patients completed qualitative questionnaires that sought to understand their experiences with the program and its overall suitability for patients being treated with eating disorders. The questionnaire data were subsequently analyzed using the thematic analysis method proposed by Braun and Clarke (Braun & Clarke, 2006). Additionally, patients completed quantitative questionnaires at three different points during their visits—at the start of the day, before visiting the galleries, and after completing an art therapy workshop with one of the authors of the study. Specifically, Thaler et al. (2017) assessed the impact of the program on patients’ (a) self-reported mood using the Profile of Mood States scale (Lorr, McNair, & Fisher, 1982), (b), eating disorder-related preoccupations (e.g., urge to weigh oneself) using a numbered Visual Analog Scale (see also Bruce et al., 2009), and (c) body image concerns using the Body Satisfaction Scale (Slade, Dewey, Newton, Brodie, & Kiemle, 1990). Basic quantitative assessments (e.g., t-tests, Means and Standard Deviations) were conducted on patients’ responses on these three scales.
In addition to being published in peer-reviewed academic journals (e.g., Baddeley, Evans, Lajeunesse, & Legari, 2017; Lachapelle, Keenlyside, & Dousenard, 2016; Thaler et al., 2017), research and evaluation findings were also reported as being disseminated during annual meeting of various professional associations (e.g., De Broux-Leduc et al., 2017a; De Broux-Leduc et al., 2017b; Fortune & Mendoza, 2017;) as well as published as theses, dissertations, and book chapters (e.g., Legari, Lajeunesse, & Giroux, in press). Findings were likewise shared both informally during staff meetings as well as via formal reports (e.g., annual reports, program reports, evaluation reports). Thomas Bastien explained that his department prepared such reports on a consistent basis:

Even if [a funder or committee] doesn’t ask for one, we give them a report. For example, our wellness programs are funded by the Rossy Family Foundation, and this year we’ve done a huge report . . . so this is a way to let them know what is going on in the museum thanks to their money.

*Satisfaction with visitor studies.* Overall, interviewees at the MMFA expressed being fairly satisfied with the number, quality, and frequency of the visitor research they conducted. “We are leading the path,” said Thomas Bastien referring to the applied research and evaluation studies undertaken under his leadership. One member of his department did, however, express wanting to see the museum use more arts-based research methodologies and representations of data in future studies of its art therapy programs. “I think there is a bridge that can be made between the user experience and the research we are doing at the museum that has a creative voice to it,” said Stephen Legari. With regards to the museum’s capacity for market research, Jean-Sébastien Bélanger noted:
Right now, I think that what we are doing is fine. I have to say that most museums are not even at our level because they lack either the competence or a computerized data base. The only museum that we really benchmark against in Canada is the Art Gallery of Ontario.

**Evaluation capacity building.** While some interviewees reported having attended conference presentations on various topics related to visitor studies at annual meetings of professional associations such as the Visitor Studies Association, the MMFA was not actively engaged in building the evaluative capacities of its staff. The lack of engagement in direct ECB was perhaps owing to the fact that a partnership approach to research and evaluation—one that relies on the research expertise of external experts—negates the need to invest in developing substantive evaluative knowledge and skills internally.

While this reasoning appeared to hold true in the case of the MMFA, a couple of educational programs officers from the Education and Wellness department nevertheless expressed interest in further building their capacities around evaluation. Specifically, Marilyn Lajeunesse, Educational Programmes Officer–Adults and Community Groups, described wanting to learn how to use Microsoft Excel to calculate descriptive statistics for her datasets while Stephen Legari indicated an interest in learning more about participatory evaluation approaches. He explained:

I feel under-qualified to provide more empowered-oriented opportunities for participant feedback. I know people that are experts at this... participatory evaluation it’s called. But I want to learn, and I want to learn how to apply those approaches specific to this context because I think we can all learn from it.
In future, the MMFA might therefore consider undertaking training efforts in the foregoing areas as a complement to its partnership approach to visitor studies.

**Evaluation challenges and needs.** Although the MMFA had a high capacity for both applied research and program evaluation, and market research, it nevertheless faced a number of challenges to their implementation. For example, several interviewees, expressed being challenged by a lack of time that they described as preventing them from engaging in various research and evaluation activities. Speaking to this limitation, Marilyn Lajeunesse said:

> The main challenge is finding the time. It really is that. I did a whole series of workshops this spring with students at the medical faculty at McGill, and I’ve been wanting to analyze the evaluations they filled out at the end of each session and send that information to the doctors at the faculty, but I just haven’t had the time to do it, you know?

An outdated computerized data base was further described by interviewees across departments as making the task of evaluation more difficult than it needed to be. Jean-Sébastien Bélanger described why the MMFA’s database posed a major challenge for him in his work and identified a new one as a pressing need:

> The database that we have right now is getting to be a problem. It’s very old, and it’s not a good database anymore. While it collects data, there’s no data analysis or reporting function built in. So, I have to extract and analyze my own data and do my own reports. I wish that it could be replaced with a really good Customer Network Management system. Then I could hire a real data analyst to work under me. As it stands, our database is not sophisticated enough for us to really maximize that sort of employee.
Another frequently mentioned challenge was a lack of financial resources for evaluation, which was attributed to the following two factors (a) The MMFA’s donors and program funders do not provide the museum with dedicated evaluation funds nor do they allow it to direct a portion of a total grant value towards evaluation and (b) Canadian nonprofits are not deemed eligible to administer grants and awards from Canadian federal and provincial research funding agencies (e.g., the Fonds de recherche du Québec, the Social Science and Humanities Research Council of Canada). “The financial support we receive for programs does not cover research expenses, and because we are not a university, we can’t apply for grant funding from the Fonds de recherche du Québec,” explained Thomas Bastien.

As the following comment by Pascale Chassé illustrated, while a dearth of evaluation funds had some negative consequences on the MMFA’s ability to conduct visitor studies the limitation has also led the museum to adopt the strategy of partnering with external experts in research and evaluation:

The last large-scale survey I did was in 2014. Last year I wanted to conduct one as well, but I didn’t have the budget for it. This year I will do another one, but I’m doing it with some graduate students from the École Polytechnique. So, we developed the questionnaire with them, and they will do all the data management and analysis after the survey is completed by our visitors.

Canadian donors and funders could therefore play a greater role in supporting visitor studies in art museums by providing them with the financial resources necessary to undertake studies. Likewise, Canadian federal and provincial research funding agencies could revisit their applicant eligibility criteria to make it easier for art museums with a research agenda to apply for grants.
Concluding comments. Over the last two decades, visitor studies have become essential components of the MMFA’s operations. The MMFA had a well-developed capacity for all three domains of visitor studies inquiry (market research, scholarly research, and evaluation) and was a leader in the sector in terms of applied research and evaluation as was reflected in the large number of studies it undertook and subsequently co-published in any given year. Its high capacity to both do and use applied research and evaluation could be attributed to a number of factors. First, such studies had the support of the highest levels of the museum as well as numerous voluntary advisory committees. Second, the MMFA had adopted a partnership approach to research and evaluation that allowed the museum to overcome an important limitation of resource scarcity (i.e., a lack of financial resources for evaluation). The strategy of partnering with external experts in research and evaluation offers a new vision of program evaluation practice and capacity for art museums. The MMFA therefore can serve as a model for other Canadian art museums interested in increasing their capacities around evaluation.

Case Portrait 2: The Art Gallery of Ontario

Introduction and study context. Occupying 45,000 square metres of physical space, the Art Gallery of Ontario (AGO), located in downtown Toronto, is Canada’s third largest art museum, next to the Montreal Museum of Fine Arts and the National Gallery of Canada, and one of the largest in North America. The AGO’s mission is “to bring people together with art to see, experience, and understand the world in new ways” (AGO, 2019b, para 3). Embedded in the museum’s vision and values is a focus on providing an exceptional visitor experience. As will be discussed in greater detail in the section below, this focus undergirds the institution’s commitment to visitor-centred decision making based on ongoing surveys and other related research.
The AGO’s permanent collection holds close to 95,000 works spanning numerous artistic movements and areas of art history from the first century to the present day. Significant collections include the world’s largest collection of Canadian art, a collection of paintings and sculptures made in Europe between 1000 and 1900, an African art collection containing work created by various peoples from south of the Sahara Desert during the 1800s and 1900s, and a modern collection that encompasses European and American art from 1900 to the 1960s. The AGO likewise has an impressive collection of photographs that spans the history of photography from the 1840s to the present day and a prints and drawings collection comprising over 20,000 works on paper dating back to the 1400s (AGO, 2019c). In addition to numerous exhibition galleries, the AGO also houses the Edward P. Taylor Library and Archives and the Marvin Gelber Print and Drawing Center.

In the last fiscal year, the AGO reported welcoming more than 965,000 visitors including 46,000 students who participated in the museum’s school programs (AGO, 2017a). It likewise boasted a VIP base numbering of close to 100,010 members, making it the second-highest ranked art museum in Canada for both attendance and membership (AGO, 2017a). As a not-for-profit organization, the AGO relies on a balance of public and private funds to operate. Specifically, 32% of its annual operating budget comes from such sources as admission ticket and food and beverage sales while another 20% comes from donations and membership fees and a further 31% comes from government grants (i.e., the Ontario Ministry of Tourism, Culture and Sport; the City of Toronto; the Canada Council for the Arts; and the Department of Canadian Heritage) (AGO, 2017a).

**Culture.** The AGO has transformed itself in significant ways since its much publicized 2008 Frank Gehry architectural re-design. Interviewees attributed the shift in the institution’s
culture towards enthusiastically embracing both a renewed commitment to life-long-learning and
data-driven visitor-centred decision making as important tools in this transformation.

*Art museum education.* The 2011 opening of the Weston Family Learning Centre—a
35,000 square foot facility dedicated to art instruction—dramatically increased the AGO’s
educational programming capacity for students, teachers, and community members. The centre
includes 6,000 square feet of studio space, three seminar rooms, a community gallery, a hands-on
centre for young children, an educational commons to welcome school groups and host events, a
youth centre for young adults, and an artist-in-residence studio (AGO, 2019d). The Weston
Family Learning Centre offers a full roster of programs for all ages. Programs include both
courses and single day workshops for adults, children, and youth, drop-in programing for youth
ages 14–15, camps, weekly Family Sunday activities for children and their caregivers,
professional development offerings for teachers, hands-on artmaking workshops for school
groups, and studio residencies for artists.

Alongside those programs facilitated at the Weston Family Learning Centre, the AGO offers one- and two-week camps for children ages 5–13, a year-long Youth Council program for youth ages 15–24, a host of film screenings, performances, art parties, and tours and talks in addition to a range of accessibility programming. Of note is one such program—the *Seniors Arts Engagement Program*—designed for older adults living in Long Term Care facilities or group residencies. This program provides participants with both social and creative experiences at the museum beginning with a tour of the AGO’s collection. After the tour, participants enjoy a catered lunch before being engaged in a hands-on therapeutic artmaking workshop (AGO, 2019e).
**Visitor studies.** The AGO has a particularly high capacity for two of the three inquiry domains of museum visitor studies—market research and evaluation (i.e., exhibition evaluation). Having formally instituted a pan-institutional visitor research and evaluation program back in 2008, Judy Koke (former Richard and Elizabeth Currie Chief of Public Programming and Learning) is credited with fostering an evaluative thinking culture within the organization, leading to the AGO’s high capacity in the foregoing domains. During her interview, Shiralee Hudson (Lead Interpretive Planner–Public Programming and Learning) noted that, over the last decade, the AGO has “built up a real appetite for visitor studies.” This appetite is reflected in the fact that, at the time this research was conducted, the AGO was the only art museum in Canada with a visitor studies department that employed internal evaluation staff in a permanent capacity. In addition to relying on such staff to conduct a wide range of studies, the museum likewise partnered with external experts (e.g., researchers, graduate students, consultants) working at local universities, marketing firms, and both nonprofit and municipal organizations on an as-needed basis to bolster their internal evaluative capacity. Despite being quite active in evaluating its exhibition projects and marketing campaigns, Judy Koke characterized the AGO’s capacity for program evaluation as still developing:

> We’re not very strong yet at evaluating our educational and public programs . . . but it’s certainly something we’re trying to do and we have done. It’s not like we don’t do it. We just don’t do it systematically, and we don’t do it as well as we should.

**Demand for visitor studies.** Visitor studies at the AGO are driven by both internal and external requirements. While interviewees identified different reasons for undertaking market research, scholarly research, and evaluation depending on their role and the department they worked for, five main interconnected drivers of visitor studies emerged from their comments.
First, leadership interest in and commitment to ongoing data collection as a basis for evidence-informed decision making was identified by interviewees as a principal driving force behind expectations for visitor studies at the AGO.

A related driver was linked to the museum’s institutional values. Specifically, the AGO’s belief that the experiences of its visitors are paramount (AGO, 2019b) was reported by interviewees as being a significant motivator for its market research and exhibition evaluation activities. “If you want to put visitors more squarely at the centre of the museum experience, you need to start with the visitor rather than start with the art, and the only practical way to do that is through visitor research and evaluation,” Judy Koke explained.

The third driver was related to the information needs of senior managers and program officers—particularly those working in the AGO’s Membership and Marketing department. Nympha Patel (Director of Membership and Database Marketing Communications and Brand) explained why data were so intrinsic to her work:

It’s really about trying to acquire new members and to keep them and then all the while measuring their lifecycle with us, and that’s through effective direct response database marketing communications strategies, and in order to do that data is so important.

Program monitoring, tailoring, and improvement was reported by interviewees as another key reason behind the demand for visitor studies information at the AGO. Referring to educational program evaluation activities at the museum, Melissa Smith (Coordinator of the Gallery Guide Program, Adult Education Officer Program and Access to Art Program) noted, “[The data we collect] help inform how we shift or respond. So, it really informs programming first and foremost.” Finally, the last driving force behind visitor studies at the AGO was both the internal and external demand for accountability. “Our development department need to report
back to either the government or our funders,” Nympha Patel explained. With few exceptions (see the Visitor Studies Activity section), interviewees reported primarily collecting output information (e.g., demographic information, participant satisfaction, amount of revenue generated through ticket sales, frequency of audio guide usage, etc.), regardless of the domain of visitor studies inquiry that they engaged in.

**Resources and supports.** As mentioned earlier, the AGO was the only art museum in Canada with a dedicated visitor studies unit—the Interpretation and Visitor Research division. Operating as one of the four divisions comprising the Public Programming and Learning department, the division primarily undertook evaluations of the museum’s exhibitions programs and was headed by Keri Ryan (Director, Interpretation and Visitor Research) whose responsibility it was to lead the evaluation process in collaboration with staff working across other departments in the institution (e.g., marketing, development, fundraising, exhibitions, and communications). In addition to Keri Ryan, the division employed three full-time interpretive planners—20% of whose time was devoted to visitor research—one nearly full-time audience researcher, and two part-time data collectors. Paola Poletto (Manager, Studio and Group Learning) described what she thought were the benefits of the AGO having institutionalized its evaluation function in this manner:

I think that having a visitor research department with a team of researchers is ideal because it allows the gallery to make connections to other departments in a way that allows for a birds-eye view of where the gaps are and where the highlights are.

While the Interpretation and Visitor Research division did not tend to hire external consultants to provide additional evaluation services because the AGO was a unionized environment, it did have an ongoing partnership with the Master of Museum Studies program
housed within the Faculty of Information at the University of Toronto. “One of the courses they teach in that program is an evaluation course, and students enrolled in the course do evaluation projects here,” Judy Koke explained.

Unlike exhibition evaluation, the responsibility of evaluating the AGO’s educational programs fell to program staff working in one of the three remaining divisions of the gallery’s Public Programming and Learning department—adult programs, young audiences, and special projects. As Paola Poletto pointed out, such staff were often supported in their evaluative work by colleagues in other departments. She noted, “I get a lot of feedback from our contact centre.” Melissa Smith seconded this notion when she said, “Our food and beverage team as a follow up to any event they host will send out a survey, and we actually get a lot of feedback about our programs through that.”

In addition to receiving evaluation support internally, program staff also reported that the AGO had recently conducted its first-ever applied research/program evaluation project in partnership with a university-based expert. Specifically, staff responsible for adult programming collaborated with Dr. Lucia Gagliese, Associate Professor in the School of Kinesiology and Health Science at York University, and her graduate students to evaluate the gallery’s Senior Arts Engagement program (AGO, 2019e). Melissa Smith explained that the strategy of working with “our partners who are deeply entrenched in research and academia” enabled the AGO to carry out a study that was more “robust” than it would have been if conducted internally. This finding aligns with observations made by Owen and Visscher (2015) who argued that museum-university partnerships brought “substantial academic expertise” (p. 73) to the work of museum educators and contributed towards building their capacities for evaluation.
While the Public Programming and Learning department undertook exhibition and educational program evaluation at the AGO, the Membership and Marketing and Visitor Services departments were responsible for leading market and other related forms of audience research and evaluation at the gallery. The heads of these departments were supported in their work not only by subordinate staff who, as Nympha Patel explained, were “all supposed to be collecting data” but also by peers working in other departments. “I work really closely with my colleague in the Visitor Experience department,” she noted. In addition to collecting data internally, the Membership and Marketing department contracts out the services of an external advertising agency to assist them with market research studies on an as-needed basis.

The AGO’s visitor studies activities are supported by the work of a number of internal committees including the Insights Working Group—a pan-institutional group of people that meets on a quarterly basis to both discuss data and set the AGO’s visitor research agenda. Describing the Insights Working Group, Judy Koke stated:

Marketing is on it, and Membership is on it, and Visitor Services is on it, and Education is on it, and just last week I sat down with that same group of people and said, “Okay, so what do we anticipate we’re going to want to do research on? What are the big institutional questions we’re trying to collectively answer?”

Since instituting this particular committee, she further observed that visitor studies at the AGO have “evolved to become much less of a siloed process and much more of a shared process.”

Across the foregoing museum departments, staff reported using differing computerized databases, software, document sharing sites, information management tools, and marketing automation platforms to collect, manage, analyze, and share information. Melissa Smith reported using both the AGO’s volunteer information and ticketing systems to track participant numbers
and activity usage. Shiralee Hudson relied on an internal file sharing site to share findings from summative exhibition evaluations with other staff members across the institution. Nympha Patel reported using Structured Query Language (SQL) and Microsoft Excel to both manage data and calculate descriptive statistics for her data sets. She likewise used Mailchimp to track statistics on email open and click through rates and both the PRIZM5 and PRIZM5 Spectra segmentation systems to produce visitor profiles according to shared demographic and behavioral information.

Additionally, the AGO’s Information Technology department was in the process of developing an in-house computerized database system to help staff integrate their multifarious data points. Nympha Patel described its functions as follows:

What we’re trying to do is internally is create this data lake or warehouse where we can input all our different data sources, and then what we want to do is overlay some kind of data visualization tool over it so that way we can arm ourselves with the right tools to help us not only know what’s happening, but also help us make decisions from both a managerial and strategic standpoint. It will also be able to produce dashboard reports. So, that’s in the works now.

**Capacity to do visitor studies.** The Public Programming and Learning Department’s significant capacity for exhibition evaluation was owed entirely to the knowledge and skills of both its director and those staff members employed within the Interpretation and Visitor Research division. Specifically, Judy Koke had completed coursework for a doctorate in Educational Leadership at Creighton University and had considerable visitor studies and evaluation experience. This included having worked as both the Manager of Visitor Research at the Denver Museum of Nature and Science, and a Senior Researcher at the Institute for Learning
and Innovation in Portland. Her experience likewise included having taught graduate level courses in visitor research and evaluation at George Washington University and the University of Toronto. Additionally, Keri Ryan and the three interpretive planners (e.g., Shiralee Hudson), one audience researcher, and two data collectors who worked under her all held graduate degrees in Museum Studies from the University of Toronto. “Most of the staff that I hire to work with me tend to come from the Museum Studies program where they teach this visitor research course just so that I know they have the theory side of things,” she explained.

In contrast to the Interpretation and Visitor Research division, staff who worked in one of the three other divisions of the AGO’s Public Programming and Learning department (i.e., adult programs, young audiences, and special projects) reported having come to their respective positions from a wider range of academic disciplines. These included art history, studio art, dance, and arts management in addition to museum studies. As a result, while some such staff possessed formal training in visitor research and evaluation, most learned various informal program evaluation techniques on the job. “Some of those tactics I learned from my colleagues by going to conferences and hearing about best practices and things that worked,” Paola Poletto noted.

Market research at the AGO, meanwhile, was undertaken by staff with backgrounds in both marketing and customer relations management. Such staff reported either receiving formal training in this domain of museum visitor studies as part of their professional preparation or pursuing additional certification while in-service (see the ECB section). As mentioned earlier, they likewise relied on the expertise of outside partners (i.e., an advertising agency) to bolster their own already well-developed internal capacities for market research.
Capacity to use visitor studies. Not unlike the Montreal Museum of Fine Arts, the AGO used the findings of the museum visitor studies it conducted in a number different ways depending on the department in question. Within the Interpretation and Visitor Research division of the Public Programming and Learning department, interviewees reported using the results of the exhibition evaluation studies they conducted to make summative judgments about an exhibition program’s effectiveness, better understand visitors’ experiences and pathways through gallery spaces, report to the museum’s governance and funders, and inform future exhibition programming. Speaking to the latter use of findings, Shiralee Hudson noted, “We did a video talk-back station in our Ei WeiWei exhibition, and we had a lot of learnings from that, and we fed those learnings into another talk-back station in the Basquiat exhibition.” The findings of such studies were likewise used to contribute to the knowledge base both within and outside of the institution. For example, in their 2017 manuscript, Koke and Ryan set out different methods (e.g., forming visitor panels and exhibition advisory groups, inviting community members to sit on exhibition teams, embracing participatory design techniques) for gathering public input and integrating that input into the exhibition program development process.

Program staff working within the three remaining divisions of the Public Programming and Learning department (i.e., adult programs, young audiences, and special projects) likewise described using evaluation findings to contribute to the generation of new knowledge in fields such as art museum education and the health sciences. Specifically, at the time this research was conducted, Melissa Smith was in the process of co-authoring a paper with Dr. Lucia Gagliese, Associate Professor in the School of Kinesiology and Health Science at York University, and other members of her research team that reported on the impact of the AGO’s Senior Arts Engagement program.
Further to these examples of the conceptual use of findings, interviewees also noted using evaluation findings for a number of other parallel purposes. Specifically, they used evaluation knowledge to make staffing decisions; identify whether they were meeting institutional goals; test differing program models, determine which pilot projects should be rolled out as ongoing partnerships; develop new programs; make summative judgments about a program’s overall merit and worth, and report to funders and the museum’s leadership alike.

On the membership side of the Membership and Marketing departments’ operations, customer data were reported as being primarily used to acquire new members and, once acquired, to manage their relationship with the AGO through targeted communications strategies. Such communications were aimed at onboarding new members, driving them to visit the museum, and, ultimately, renewing their memberships. On the marketing side meanwhile, interviewees reported using market research data to better understand various audience segments, drive ticket sales through the development of effective marketing campaigns aimed at those audiences, and track the effectiveness of such campaigns. Market research was also undertaken to gauge potential visitors’ understandings of and interest in a proposed exhibit, and the data collected through these studies were subsequently used to, among other things, inform how the AGO’s exhibitions were both named and marketed.

The data gathered by the Membership and Marketing department were additionally reported as being used for persuasive purposes—namely to educate the institution about the value and utility of visitor research and evaluation. As Nympha Patel noted:

I think culturally it’s also a bit of proving the benefit of using data, the value of it because it’s new to this kind of world. So, for me it’s also trying to advocate the value of why good data is so important.
Visitor studies activities. As was previously mentioned, the AGO primarily engaged in two of the three domains of visitor studies inquiry—exhibition evaluation and market research. To a lesser extent, the museum also conducted applied research and program evaluation studies focused on both monitoring and examining the impacts of its educational programs. The nature of the activity occurring within each of these domains of visitor studies inquiry is described in detail below.

Exhibition evaluation. Interviewees working in the Interpretation and Visitor Research division of the Public Programming and Learning department reported primarily undertaking judgment-oriented summative evaluations aimed at understanding and measuring the success of the AGO’s temporary ticketed exhibitions against stated goals and objectives. In addition to conducting such studies on a routine basis, these staff also conducted periodic evaluations of various other types of exhibitions. For example, at the time this research was conducted, the Interpretation and Visitor Research division had just completed a major evaluation focused on examining the performance of the AGO’s European Collection following a recent re-installation. The division likewise reported enlisting students from the University of Toronto’s Master of Museum Studies program in the evaluation of smaller exhibitions that its research budget did not cover. Keri Ryan explained:

We’ll often get them to evaluate a special exhibition we’ve done that’s not ticketed. This year we used them to do some research on the [Before and after the Horizon: Anishinaabe Artists of the Great Lakes] exhibition because I didn’t have the budget to have my data collectors look at it. So, we pick projects that, as the interpretive planning team, we really wish we could evaluate because it’s something that’s different or a really great opportunity. We’ll also get them to do ‘pet projects.’ Sometimes they should be the
projects that we’re actually doing the evaluation on in the institution, but just don’t have the resources for.

Overall, the information that the Interpretation and Visitor Research division collected was output-based in nature. Specifically, interviewees responsible for exhibition evaluation reported collecting varied visitor sociodemographic data (e.g., age, gender, ethnicity, location, type of client), behavioral data (e.g., movement patterns, stay times, social interaction, use of interactive elements), and both attitudinal and experiential data (e.g., what visitors liked about an exhibition and wanted to see more of, what distracted from their museum experience). Such data were gathered using a combination of document reviews (e.g., computerized database records), direct observations (e.g., timing and tracking), online questionnaires, and face-to-face interviews. As was illustrated by the following comments made by Keri Ryan:

We do exit surveys at the end of the exhibition. So, data collectors would administer a long-form questionnaire, and we’ll often have the visitors sit down and work with us on some more in depth conversations, and then we also do an online survey that goes out to anybody who bought a ticket to an exhibition. . . We also track within our exhibitions. So, that could be a tiny track using playing cards or we will pick the fifth person and observe them in the spaces, identifying hot spots and doing a sweep rate index as far as how long someone spent in the space.

The information collected via these means was typically analyzed using either qualitative content analysis or basic quantitative analytical methods (e.g., descriptive statistics). Exhibition evaluation findings were reported as being presented in final report format and shared both formally and informally during various staff meetings. This included bi-weekly meetings of the
Interpretation and Visitor Research division in addition to post-mortem meetings held with the AGO’s leadership and exhibition teams.

*Market research.* Because data are central to their aims, those interviewees with responsibility for market research and other related forms of audience research (i.e., Marketing and Membership, Visitor Services) reported being engaged in a process of continuous data collection and analysis to better understand the AGO’s customers and optimize their marketing campaigns. Overall, such staff collected output information. This information included visitor profile data (e.g., age, gender, ethnicity, geographic location, type of client, lifestyle), behavioral data (e.g., purchasing habits, frequency of visitation, responses to messaging), and attitudinal data (e.g., motivations, preferences, levels of satisfaction). These data were collected both on-site (e.g., at the ticket kiosk, member services desk, and museum exit) and online using document reviews (e.g., of computerized database records) and face-to-face interviews. Shiralee Hudson, whose staff assisted with the latter data collection method, explained:

> There are 200 randomly selected institutional exit interviews that are performed every month . . . it’s really demographics, [Net Promoter Score] ratings—our measure of customer satisfaction—and maybe one or two additional questions about something that we might want to learn about in a particular month or quarter. Like, how many visitors have smart phones and what kind of smart phones they have or what visitors are seeing when they come. So, our data collectors will hold up a board with pictures of different exhibitions and, basically, visitors point to what they visited that day. The information collected via these various means was routinely analyzed by Membership and Marketing staff using both standard qualitative content analysis and quantitative marketing analytics techniques. As Nympha Patel pointed out, the department also
used information collected by a third-party provider in addition to using data that it collected internally:

The other data source that we use is PRIZM data through Environics Analytics. So, based on postal code data, they have created these lifestyle profiles of people. So, what we are able to do is based on your postal code; we are able to identify if you are an *Urban Digerati* for example—these are profiles of people with similar interests—and it gives us a sense of the people that visit. That’s a key data source for us. So, we use that external third-party data for a lot of things that we do, and we overlay it with the campaigns or analytics that we do in house.

Further to this, interviewees working in the Membership and Marketing department also periodically worked with an advertising agency to collect attitudinal data through focus group research aimed at gauging visitors’ interests in and understandings of the concepts behind a planned exhibition.

**Applied research and program evaluation.** Program staff working within the adult programs, young audiences, and special projects divisions of the Public Programming and Learning department reported carrying out mostly informal, improvement-oriented formative evaluations of their own programs. Not unlike their colleagues from other museums’ divisions and departments, they primarily collected information on program outputs. “We’re not in a position right now to really measure outcomes,” Keri Ryan explained. The types of data collected included sociodemographic data (e.g., age, gender, ethnicity, location, type of client, etc.), behavioral data (e.g., stay times, repeat visitation, social interaction, number of Facebook shares), and both attitudinal data (e.g., motivations, levels of satisfaction) and attendance statistics.
As Paola Poletto noted, for revenue-driven programs interviewees also reported collecting registration statistics to determine whether they were meeting revenue targets, “We have budgetary targets that we are striving to meet paired with a set number of registrations. So, we measure certain programs on successful retention and on uptake.” For the museum’s youth programming, additional measures of success included both whether youth participating in the AGO’s Free After Three program were accepted in to the Youth Council and whether Youth Council members later went on to pursue post-secondary education in the visual arts in addition to the caliber of visual artist the museum could attract to work with youth on diverse collaborations.

Program staff reported collecting output data using a combination of document reviews (e.g., computerized database records, visitor comment cards), direct observations, face-to-face interviews, both on-site and online questionnaires, and informal conversations. Melissa Smith described one such method:

For one of our Access to Art exhibitions, we wanted to have people vote on what art meant in their lives. So, based on John Falk and Lynn Dierking’s [1992, 2000] contextual model of learning in museums, we had visitors vote by dropping these disks into different plastic frames. It was a neat participatory way of collecting thoughts and responses. The data that were collected via the foregoing means were reported as being analyzed using either standard qualitative content analysis or basic quantitative analytical methods enabled by spreadsheet software such as Microsoft Excel. The findings of program evaluation studies were both disseminated in the AGO’s annual report and made available to the Development department to share with those donors and funders that required them.
As was previously mentioned, in addition to undertaking output-based formative evaluations of their programs themselves, interviewees also described having recently partnered with external experts (i.e., City of Toronto Long-Term Homes and Services staff, graduate students, a university professor) to conduct a systematic, judgment-oriented summative study of the AGO’s Senior Arts Engagement program. In this case, interviewees described measuring longer term outcomes such as happiness and quality of life using a combination of direct observation, face-to-face interviews, and online questionnaires. A co-authored paper reporting the findings of this study was in the process of being written at the time this research was conducted.

**Satisfaction with visitor studies.** Overall, interviewees working within the Interpretation and Visitor Research division of Interpretation and Public Programming department as well as the Membership and Marketing department expressed being fairly satisfied with the number, quality, and frequency of the exhibition evaluation and market research studies they conducted. “We’re leading anyway. We’re [collecting and analyzing customer data] every day in the work that we do,” remarked Nympha Patel referring to the customer relationship management strategies having been undertaken at the AGO under her leadership. This said, interviewees with responsibility for both domains of visitor studies inquiry, were also actively trying to refine their evaluation practices. For example, while Shiralee Hudson was in the process of trying to standardize how her team collected data, Nympha Patel was working on creating a common reporting template that could be rolled out across the institution.

Unlike their colleagues with responsibility for market research and exhibition evaluation, program staff working in of the adult programs, young audiences, and special projects divisions of the Interpretation and Public Programming department expressed being interested in
improving program evaluation at the AGO. “It’s really important for us to be doing more and better program evaluation,” said Judy Koke referring to her goals for the museum’s future visitor studies agenda. Keri Ryan added:

Numbers and anecdotes—that’s the only way we can tell a story right now about the impact that we’re having so it will be interesting to see if we can shift towards really measuring program outcomes.

**Evaluation capacity building.** While interviewees noted participating in indirect Evaluation Capacity Building (ECB) activities (e.g., attending conference presentations on topics related to evaluation at annual meetings of the Visitor Studies Society) regardless of the division or department they worked for, only the Directors of the Membership and Marketing and Interpretation and Visitor Research division of the Interpretation and Public Programming department reported also being engaged in formal direct ECB activities. Keri Ryan, for instance, had just completed an additional certificate in visitor research and evaluation at the time this study was conducted. Referring to her own in-service professional development around market research, Nympha Patel noted:

I have taken a lot of additional certification. I work a lot through the Canadian Marketing Association (CMA) for direct marketing and customer insights and analytics, and I’m also on the Not-For-Profit council on the CMA. I’m also part of a number of committees so I think I’ve sort of inserted myself in the membership world not just within Canada, but also in North America and internationally.

**Evaluation challenges and needs.** Although the AGO had a high capacity for two of the three domains of visitor studies inquiry (i.e., exhibition evaluation and market research), interviewees noted that engaging in visitor studies was not free of certain challenges. Across
departments interviewees expressed being challenged by a lack of time which Judy Koke attributed to the fact that “in museums everybody is doing two and a half jobs.” Specifically, time pressures were described as preventing interviewees from (a) engaging in evaluation in the first place, (b) analyzing data that had already been collected, (c) investing in training research staff, (d) developing necessary evaluation tools, (e) involving various stakeholders in the evaluation process, and (f) measuring outcomes. This finding reinforces Cutt and Murray’s (2000) and Juillet, Andrew, Aubry, and Mrenica’s (2001) observations that even personnel trained in evaluation are faced with limited available time for engaging in evaluation activities.

Another challenge that interviewees reported facing regardless of the department they worked for, was insufficient horizontal (i.e., across departments) and vertical (i.e., at all levels of the organization) coordination of visitor studies efforts across the museum. This challenge was ascribed to the fact that the AGO was a complex organization that engaged a wide variety of staff in one form of visitor study or another. “I’m trying to get my head around all the different types of visitor research that we’re doing here at the AGO because there’s lots, and I’m just seeing a little bit through the exhibitions window,” Shiralee Hudson explained. She therefore saw a need for the creation of an institution-wide visitor studies calendar that would allow the AGO to have a better sense of the scope of their research and evaluation activities so that these could be approached more strategically. Similarly, Nympha Patel described thinking that the in-house computerized database system that was currently under development would help the AGO address this particular challenge by integrating all of their data sources into a central platform that would provide staff with “a better view of what is actually happening overall from an organizational perspective.”
While interviewees across departments said that they could benefit from additional financial resources, those with responsibility for program evaluation reported being particularly challenged by a lack of dedicated evaluation funds. Unlike their colleagues with responsibility for exhibition evaluation and market research, these interviewees did not receive evaluation funding from either their institution or their funders. Referring to the dearth of external support for program evaluation as well as the optics of using a portion of a total program grant value for evaluation purposes, Judy Koke noted:

The Institute for Museum and Library Services [in the US] requires ten percent of any grant proposal to be dedicated to evaluation. So, program evaluation is really healthily funded in the States. It’s not funded here. When you spend money on evaluation, it’s seen as taking money away from program delivery, and we’re already not delivering as much programming as we would like and are already constrained by resources.

She further gave the following explanation for the lack of availability of internal program evaluation funding at the AGO:

When I was hired, it was to integrate visitor voices into exhibition or experience design. It wasn’t to evaluate whether or not our school programs were making a difference. And, if you ask most people what our biggest product is, they will say, “The exhibitions.” So, it’s perhaps unsurprising that that’s where our resources are first allocated.

A similar finding was observed in a 2004 study of museum education evaluation in the United States (Smithsonian Institution, 2004) where many large museums reported that exhibition evaluation still gets more support from their institution than does program evaluation. Unsurprisingly, several interviewees working in the Interpretation and Public Programming department identified increased institutional demand for and financial support of program
evaluation as a pressing future need. For example, when asked what would enable her to undertake more program evaluations, Keri Ryan noted, “The institution would have to decide that it was something that they really wanted to do and that they were going to put some money towards it.”

Finally, those interviewees with responsibility for exhibition evaluation expressed being challenged by the fact that some curators at the AGO still held predominantly negative perceptions about exhibition evaluation, confirming Preskill’s (2011) suggestion that “evaluation is often seen by others as threatening, anxiety producing, and potentially punishing” (p. 97). Judy Koke explained:

The exhibitions people—the project managers, the designers—and the educators on a project are all very interested in evaluation and find it a useful tool, but there’s not that same appreciation in the curatorial division. There are some curators that respond very thoughtfully to the visitor research that we do, and then there are those that are quite dismissive. I think sometimes the curators in an art museum feel like we’re evaluating their idea, which makes them defensive.

Melissa Smith went on to add that negative attitudes towards evaluation prevented certain curators from modifying their practices in the face of the knowledge the Interpretation and Visitor Research division acquired. She noted, “We definitely have challenges interacting with our curators. So, even with concrete data, people kind of refute that and then sit on their laurels and perpetuate unhealthy organizational structures.”

The above comments illustrate that as much as ongoing work needs to be done to help some curators see how evaluation can inform their exhibition program development process in a positive way, Canadian funders could play a greater role in supporting program evaluation in art
museums by providing them with the financial resources to undertake studies. Likewise, the AGO’s leadership could demonstrate a commitment to a more comprehensive approach to visitor studies—one that examines the outcomes of exhibitions, marketing campaigns, and educational programs. In doing so, the AGO will be able to show a much broader picture of the museum’s institutional value than can be reported by conducting exhibition evaluations and market research alone.

Concluding comments. Over the last decade, the AGO has made significant strides in shifting its institutional culture towards embracing evidence-informed, visitor-centred decision making. As a result, the museum had a high capacity for two of the three domains of visitor studies inquiry (i.e., exhibition evaluation and market research) and was a leader in the sector in terms of exhibition evaluation as was reflected in the large number of summative studies it conducted on a routine basis. The AGO’s high capacity for this particular domain of visitor studies inquiry was attributable to two main factors. First, precluding some curators, exhibition evaluation had the support of the museum’s leadership as well as several internal working committees. Second, the AGO was the only art museum in Canada with a dedicated visitor studies department that employed qualified staff to evaluate its exhibition programs. The strategy of hiring internal evaluation staff on a permanent basis can therefore serve as a model for other Canadian art museums interested in increasing their capacities for evaluation. A lack of both internal and external demand for and, hence, financial support of program evaluation was the greatest contributing factor to the AGO’s only developing capacity to evaluate its educational and public programs.
Cross-Case Results

In addition to painting an in-depth picture of what developed capacity for evaluation looked like in both the Montreal Museum of Fine Arts (MMFA) and the Art Gallery of Ontario (AGO), another aim of this phase of the research was to identify the similarities and differences that arose across cases for the purpose of making analytic generalizations (Lincoln & Guba, 1985). My cross-case analysis revealed a number of important differences between the cases of the MMFA and the AGO. For example, the two museums differed somewhat with respect to what they valued and how they positioned themselves in their communities. The MMFA placed its educational, art therapy, and social/community programs squarely at the centre of its mission and positioned itself as a caring, humanist, and socially-engaged museum. The AGO, meanwhile, put greater emphasis on bringing people together with art and providing their visitors with an exceptional visitor experience. Despite their differences, both museums shared the following key similarities that were both descriptive of and accounted for their overall levels of organizational evaluation capacity (a) significant capacity to do evaluation, (b) ongoing pan-institutional engagement with a wide range of research and evaluation studies, (c) the presence of an evaluation minded-culture, and (d) dedicated resources and supports. These similarities are both described in greater detail and discussed below.

Significant capacity to do evaluation. Both the MMFA and the AGO possessed significant capacities to conduct research and evaluation studies. Each museum engaged in all three domains of visitor studies inquiry (i.e., scholarly research, market research, and evaluation) to different extents. This said, the MMFA emerged as a leader in the sector in terms of applied research and program evaluation while the AGO had a particularly high capacity for exhibition evaluation. Although both museums relied on a combination of internal expertise and
partnerships with experts, the AGO’s considerable capacity for exhibition evaluation was owed entirely to the knowledge and skills of those staff members employed within its Interpretation and Visitor Research division who were encouraged to continue developing their expertise on the job through participation in both direct and indirect ECB activities. The MMFA’s well-developed capacity for evaluation and applied research, meanwhile, was owed less to the knowledge and skills of internal staff and more to the museum’s broad-based partnership approach to evaluation—an approach that has also been referred to as coevaluation (Gray, 1998). The case of the MMFA highlights the importance of such an approach for those art museums with limited in-house technical expertise.

**Ongoing pan-institutional engagement with a wide range of research and evaluation activities.** An indicator of both the MMFA’s and AGO’s high capacities for evaluation was their ongoing engagement with various forms of visitor studies inquiry even when it was not mandated by funders. At both museums, staff members with responsibility for market research reported being engaged in a process of continuous data collection and analysis to better understand visitors and optimize their marketing campaigns. Similarly, program staff at both the MMFA and the AGO also conducted informal formative studies as part of a continuous cycle of improvement. Further to this, staff members with responsibility for exhibition evaluation at the AGO reported evaluating the museum’s temporary ticketed exhibitions as a matter of routine while, in partnership with various external experts, program staff at the MMFA evaluated an impressive number of programs each year. Specifically, at the time this research was conducted, the MMFA was involved in no fewer than 32 collaborative research/evaluation studies—five of which were focused on its accessibility and inclusion programs, 13 of which were focused on its
educational and public programs, and another 14 that focused on the museum’s art therapy and health and well-being programs.

**The presence of an evaluation-minded culture.** Both the MMFA and the AGO had well-established evaluation cultures. The presence of an evaluation habit of mind (Katz, Sutherland, & Earl, 2002) at each museum was evidenced through, among other things, leadership demand for evaluation, the adoption of research and evaluation as a core organizational value, and a view of evaluation as an important pathway towards the achievement of fundamental institutional values. Specifically, long-term leadership interest in and support of two or more domains of visitor studies inquiry were a key drivers of research and evaluation activities at both museums. Both the MMFA and the AGO had also spent the last one to two decades actively developing cultures that valued visitor studies. As a result, research and evaluation were incorporated both at all levels of the institution and across various departments through appropriate structures at both museums. Furthermore, in both cases, research and evaluation were described as integral to achieving certain fundamental institutional values. At the AGO visitor studies were perceived as an avenue to providing an exceptional visitor experience while at the MMFA research and evaluation were key tools that were used in the service of justice (i.e., to promote social cohesion and individual health and well-being).

**Dedicated resources and supports.** The cases of the MMFA and the AGO also allow us to see the *resources and supports* factor in my conceptual framework at work. In addition to being advocated by leadership, research and evaluation at both museums likewise benefited from support in the form of time, human, and technological resources. For example, while staff members with responsibility for exhibition evaluation at the AGO were expected to devote 20% of their time to this task, program staff at the MMFA were frequently freed up from their regular
job duties to engage in applied research and program evaluation studies alongside various external experts. Staff members with responsibility for visitor studies at both museums also received varying degrees of support with data collection, analysis, and reporting from staff members working both within and outside of their own departments. They were further supported by the work of several volunteer committees. Specifically, at the AGO visitor studies had the support of internal committees such as the Insights Working Group—a pan-institutional group of people which meets on a quarterly basis to both discuss data and set the AGO’s visitor research agenda. The MMFA’s applied research and evaluation activities, on the other hand, were supported by the work of several advisory committees whose members shared their expertise on best practices in their respective fields, evaluated the merits of potential lines of inquiry, and provided a vision for the museum’s overall research agenda in addition to suggestions for promising partnerships. Moreover, in both cases, staff members with responsibility for visitor studies had access to both sophisticated computerized data bases that integrated multiple streams of data across various departments on a continuous basis and various software, document sharing sites, information management tools, and marketing automation platforms which assisted them in collecting, analyzing, managing, and sharing information.

Concluding comments. Many of the ideas explored in the cross-case analysis above (e.g., leadership backing, a culture of inquiry, evaluative capability, appropriate supports, etc.) seem to fit with the broader literature on both evaluation capacity and mainstreaming\(^{19}\) (e.g., Bourgeois & Cousins’, 2013; Cousins et al., 2008; Gullickson, 2010, 2017; Taylor-Ritzler et al., 2013). Thus, both the MMFA and AGO can be said to have had high capacities for research and evaluation because they had successfully mainstreamed them into their daily operations. My

\(^{19}\) Sanders (2003), defines evaluation mainstreaming as “the process of making evaluation an integral part of an organization’s everyday operations (p. 3).
cross-case analysis also revealed several strategies (two of which have already been articulated by Gullickson [2017], as possible paths to the ultimate goal of mainstreaming) that art museums might employ to bolster their evaluative capacities. These included employing qualified staff to undertake studies internally, partnering with external experts in research and evaluation, forming evaluation committees, and engaging staff in direct ECB. While these strategies are in no way mutually exclusive and are likely to be that much more effective when used in combination, the cases of the MMFA and AGO suggest that developed EC in art museums may be reached from different starting points and in several different ways. Moreover, depending how closely their own situations match those of the case studies, these broad strategies may also be transferrable to other contexts such as arts-based nonprofit organizations and other museum types.
Chapter 6
Summary, Discussion, and Implications

I begin this chapter by both revisiting my dissertation problem, purpose, and questions and by providing an overview of the study. Next, I present a summary of the key findings by research question from both Phase One and Phase Two of this study and discuss these findings in relation to the broader literature. This discussion is followed by the implications of this research for theory, policy, practice, and future research. I conclude the chapter by briefly outlining the study’s significance.

Dissertation Purpose, Questions, and Study Design

Towards the latter part of the 20th century art, museums gained increased recognition as educational institutions (Weil, 2002; Willumson, 2007; Xanthoudaki, Sekules, & Tickle, 2003). Today, most art museums place learning at the center of what they do (Falk & Dierking, 2000). It is no longer enough, however, to simply claim that art museums are inherently educational. Rather, in what has become known as the age of accountability, governments and external funding agencies, particularly in the US and the UK, have increasingly called upon art museums to demonstrate evidence of program effectiveness (Gorman, 2007; Korn, 2007). As a result, the field has seen a rise in the number and frequency of program evaluations conducted in art museums in these countries and elsewhere over the past thirty years (Adams, 2012). Despite this trend, this important area of professional practice has not received much attention in either the museum or evaluation research literature. Consequently, little is known about what art museums are doing to evaluate their educational programs, to what extent they possess the internal capacities to conduct evaluations and use evaluation findings, what efforts they engage in to
reflect on and build this capacity, what their evaluation resource and support needs may be, or the challenges they face when attempting to evaluate programmatic impact.

Recognizing the disparity between the current status of professional practice and research into program evaluation in the context of art museums, numerous scholars have called for an in-depth examination of evaluation practices and capacities within this sector (Hein, 1994, 2005; Parsons, 2004; Sheppard, 2000, 2010; Weil, 2003; Worts, 2006). What was first needed to guide such an investigation was a conceptual framework that demarcated what evaluation capacity (EC) might look like in these unique professional settings. While several frameworks exist in the evaluation literature that describe EC in government and school settings as well as disability-related community-based organizations, and (e.g., Bourgeois & Cousins, 2013; Cheng & King, 2017; Cousins et al., 2008; Nielsen, Lemire, & Skov, 2011; Suarez-Balcazar et al., 2010; Taylor-Ritzler et al., 2013), prior to the start of this study, such a framework had yet to be delineated, even though several scholars have argued for more research that advances the evaluation field’s understandings of EC across a variety of sectors and organizational types (e.g., Cousins & Bourgeois, 2014; Taylor-Ritzler et al., 2013).

Therefore, one purpose of this study was to address the foregoing gaps in both the art museum education and evaluation literature by first developing a framework that conceptualizes the dimensions that might comprise EC in art museums and subsequently examining, across its multiple dimensions, what program evaluation practice and capacity looked like both sector-wide and in those Canadian art museums that were most active in conducting a wide range of visitor studies. A related purpose of this study was to draw out specific examples of institutional successes for the purpose of sharing them with the growing number of art museums seeking to leverage evaluation’s learning benefits to support positive change in practice (e.g., to learn about
their programs’ strengths and weaknesses in order to improve them and, ultimately, enhance museum learning). Three central research questions guided this study:

1. What are the key dimensions that describe evaluation capacity in the Canadian art museum sector?

2. How is program evaluation capacity manifested across Canada’s largest publicly owned art museums, and to what extent does this capacity vary from museum to museum?

3. How is overall organizational evaluation capacity manifested in those Canadian art museums most active in conducting a wide range of visitor studies, and which factors and conditions, according to practitioners’ self-reported understandings, enabled their capacities to develop to their current levels?

A two-phase multiple method qualitative research design was used to address the purposes of this research and answer the above research questions. Phase One involved conducting an interview study to both establish an initial knowledge base on Canadian art museum educators’ program evaluation practices and capacities and test the degree to which the literature-derived conceptual framework I developed to guide this study, presented in Chapter 2, could be considered an accurate and complete description of EC in the art museum context. Phase Two involved conducting qualitative case studies (Yin, 2014) of two art museums that, based on my interview findings, I identified as operating at the highest level of capacity for evaluation in the country: the Montreal Museum of Fine Arts and the Art Gallery of Ontario. Case study research was selected for this phase of the study to enable an in-depth exploration into what developed capacity looked like in the Canadian art museum sector, how these museums’ capacities for program evaluation compared to their capacities to conduct studies in
other domains of visitor studies inquiry (i.e., scholarly research, market research, and exhibition evaluation), and to elucidate specific examples of good practice.

**Discussion of Key Findings by Research Question**

The key findings of both phases of this study, organized by research question, are presented and discussed in the following paragraphs.

**Key findings pertaining to research question number one.** The first research question guiding this study was: *What are the key dimensions that describe evaluation capacity in the Canadian art museum sector?* Based on an analysis of existing EC frameworks (i.e., Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen et al., 2011; Suarez-Balcazar et al., 2010; Taylor-Ritzler et al., 2013) as well as the museum literature (i.e., art museum education, visitor studies), in Chapter 2 of this dissertation I presented an initial *Conceptual Framework of Evaluation Capacity in Art Museums* that articulates the key dimensions of EC as they could be observed in these unique learning settings. The validity of this framework was tested through the empirical portions of this study. In this section, following a brief review the initial framework, I first provide validity evidence for those framework dimensions conceptualized as both *influencing factors* and *capacity indicators*. I conclude the section by discussing the relationships that emerged among the framework’s individual dimensions.

**Revisiting the conceptual framework of evaluation capacity in art museums.** The initial framework was comprised of seven dimensions that I posited were essential to understanding evaluation practice and capacity in art museums. These seven dimensions are organized under two main categories in my conceptual framework. Specifically, the first category in my framework—*influencing factors*—describe the factors that impact evaluation practice and capacity in art museums in addition to the extent to which such museums are likely to engage in
evaluation capacity building (ECB) activities. This category comprises four dimensions: **context**, **culture**, **demand for evaluation**, and **resources and supports**. **Context** refers to such characteristics as an art museum’s size, mandate, organizational structure, and extant level of institutional stability versus change. **Culture** seeks to capture an art museum’s institutional and evaluation culture. **Demand for evaluation** assesses the drivers of evaluation in art museums (internal vs. external) and aims to gain insight into funders’ accountability expectations. **Resources and supports**, meanwhile, looks at the type and extent of the resources and supports that are allocated to evaluation in any given art museum.

The second category in my framework—**capacity indicators**—consists of three dimensions conceptualized as interrelated capacity indicators: **capacity to do and use program evaluation**, **program evaluation activity**, and **direct ECB**. The **capacity to do and use program evaluation** describes the degree to which those practitioners with responsibility for research evaluation in Canadian art museums possess the requisite knowledge and skills to conduct program evaluations and use the findings of such investigations for diverse purposes. **Program evaluation activity**, on the other hand, is concerned with understanding the nature and extent of the program evaluation activity occurring in art museums. Both dimensions are conceptualized as nested constructs because, in art and other museums, the evaluation of educational programs takes place within a larger body of research-based knowledge production (i.e., museum visitor studies). The final dimension in this category, **direct ECB**, examines whether art museums are actively engaged in building their staff’s evaluative capacities and, if so, the specific activities they use to do so.

**Influencing factors.** Precluding a few minor but necessary adjustments, considerable evidence was found to suggest that influencing factors such as **context**, **culture**, **demand for**
evaluation, and resources and supports can contribute to either promoting or limiting an art museum’s overall capacity for evaluation. For example, capacity was promoted when leadership championed evaluation, when a museum was characterized as being less formalized in terms of its organizational structure, and when practitioners were required to undertake evaluations by their leadership and funders alike. The availability of technology as well as training and human resource support—both from internal staff members and various external experts—likewise served to promote capacity in Canadian art museums. Conversely, a lack of funding in addition to either a lack of interest in or awareness about evaluation and its importance or outright resistance to evaluation on the part of groups or individuals not surprisingly had a marked negative influence on the uptake of evaluation across the sector.

In addition to those sub-dimensions already captured in my initial framework, an additional two sub-dimensions emerged from my analysis of the Phase One interview data. Specifically, my findings revealed that whereas an art museum’s extant level of institutional stability was a key contextual characteristic influencing capacity for evaluation, the extent to which an art museum could be classified as a learning organization emerged as an important cultural characteristic that served to either limit or promote capacity. For example, Canadian art museums’ capacities for evaluation were reported as being limited by wide-ranging organizational changes (e.g., expansion/renovation projects, strategic planning initiatives, turnover in leadership, etc.) that served to siphon considerable financial and human resources away from evaluation activities and often resulted in the loss of a museum’s primary evaluation advocate. A lack of institutional commitment to evaluation was further attributed by several educators to their museums’ broader lack of commitment to organizational learning.
These findings, together with those of others (e.g., Bourgeois and Cousins 2013; Cheng & King, 2017; Cousins et al., 2008; Taylor-Ritzler et al., 2013; Suarez-Balcazar et al., 2010) lend further credence not only to the inclusion of influencing factors in any conceptualization of EC but also, given their significant impact on an organization’s capacity for evaluation, to the need to explicitly consider cultural and contextual characteristics among them.

**Capacity indicators.** In addition to providing validity evidence for those dimensions conceptualized as influencing factors in my conceptual framework, the findings from this study likewise indicate that EC in the art museum sector can be described through two of the dimensions conceptualized as capacity indicators in my study: *capacity to do and use program evaluation* and *program evaluation activity*. While the findings of this study provided significant insight into Canadian art museums’ engagement and/or lack thereof with ECB, they also raised questions about the location of this important construct within the framework. In other words, ECB might be ill-conceived of as an *indicator* of capacity, at least in the Canadian art museum sector.

Specifically, while it was expected that a higher level of EC would be associated with an art museum’s use of effective ECB processes, this turned out not to be the case at an organizational level despite the fact that involvement in direct ECB activities had significant long-term impacts on individual participants. Rather, in the two art museums identified as operating at the highest level of capacity for evaluation in the country—the Montreal Museum of Fine Arts and the Art Gallery of Ontario—developed EC was associated with a partnership approach to evaluation and the existing evaluative knowledge and skills of staff that ECB only served to complement.

This finding may be explained, in part, by some data loss. Specifically, had the third art
museum selected for case study agreed to participate in the second phase of this research, more insight might have been gained into whether or not the engagement of staff in multi-day face-to-face training workshops with the principal of a leading visitor studies firm contributed to its overall capacity for evaluation. As it stands, two art museums participated in Phase Two instead of three, and the findings from the second phase of this study suggest that, rather than being conceptualized as a stand-alone dimension falling under the capacity indicators category of my framework, ECB may be better embedded within (a) the knowledge and skills sub-dimension falling under the capacity to do and use program evaluation dimension or (b) as a sub-dimension of the resources and supports dimension.

Relocating ECB in this way would better align my framework with the other EC frameworks that have been published to date in which the construct appears as a sub-dimension under the human resources dimension in Bourgeois and Cousins’ (2013) profile framework, a stand-alone dimension (i.e., sources of knowledge, skills and activities) in Cousins’ et al.’s (2008) conceptual framework, a sub-dimension falling under the human capital dimension in Nielsen et al.’s (2011) conceptual model, or subsumed under the competence dimension in Taylor-Ritzler et al.’s (2013) synthesis model. Clearly, further testing is needed to determine the proper location of ECB within the Conceptual Framework of Evaluation Capacity in Art Museums.

The relationship between framework dimensions. In my initial framework, I clearly laid out how the dimensions conceptualized as influencing factors related to those conceptualized as capacity indicators (i.e., the former impact the latter). I likewise made explicit the hypothesized relationship among the three dimensions conceived of as capacity indicators—capacity to do and use, evaluation practice, and direct ECB activity. For instance, a more developed capacity for
evaluation was assumed to lead to higher quality evaluation studies while engaging in evaluation activities was presumed to feed back into a museum’s capacity to do and use evaluation and lead to what Cousins, Goh, and Clark (2006) described as data valuing. What I had not anticipated, however, was the relationship that emerged among the four dimensions conceived of as influencing factors (i.e., context, culture, demand for evaluation, resources and supports).

Specifically, the Phase One results revealed that organizational structure (a contextual sub-dimension) was connected to the cultural sub-dimensions of valuing of museum education and capacity for organizational learning. In other words, contextual characteristics such as a museum’s organizational structure and mandate were found to reflect certain aspects of its institutional culture. Aspects such as the value a museum placed on education and the degree to which it demonstrated markers of organizational learning like clarity and support for mission and vision, shared leadership, and teamwork—markers that are facilitated by flat decentralized organizational structures (Goh, 2000). This finding suggests that, rather than being conceptualized as two separate dimensions falling under the influencing factors category of my framework, context and culture may be better conceptualized as a single dimension.

Further, the findings of both phases of this study clearly indicated that an art museum’s contextual and cultural characteristics were linked to the evaluation expectations of both a museum’s leadership and its funders as well as the extent of the resources and supports that they made available for evaluation. For example, Canada’s low accountability context resulted in relatively little funder demand for and hence concomitant financial support of evaluation in art museums. Similarly, a museum’s mission and values drove evaluation in the Canadian art museum sector and determined both what got evaluated (i.e., educational programs versus exhibitions) and whether or not internal resources were dedicated to the evaluation function.
Together, the above unexpected findings highlight the degree to which the framework dimensions and sub-dimensions are interdependent in practice and serve to deepen the implications for understanding EC in the art museum context.

**Key findings pertaining to research question number two.** The second research question guiding this study was: *How is program evaluation capacity manifested across Canada’s largest publicly owned art museums, and to what extent does this capacity vary from museum to museum?* The findings from the first phase of this study provided significant insight into Canadian art museums’ program evaluation practices and capacities, their level of engagement with ECB, and the factors that influenced their overall EC. Moreover, the findings of this study highlight both the challenges Canadian art museums face on their journeys towards evaluating programmatic impact and their needs for assistance around evaluation. In this section, I summarize and discuss the key findings that emerged from this phase of the research.

**The value assigned to art museum education.** Only half of the interviewees I spoke with reported thinking that education played an important role at their institutions. Thus, in half of Canada’s art museums, a lack of rigorous program evaluation activity was a symptom, at least in part, of a broader cultural issue—the systematic marginalization of educational endeavours in museums. As Luke and Knutson (2010) have argued, such art museums “need to embrace a strong and clear educational mandate, not only within education departments but institution-wide” (p. 253).

**Embracing evaluative thinking.** All the interviewees I spoke with recognized both the need for and value of evaluating their educational programs, however, evaluation as an organizational value was not necessarily evenly integrated across all departments or levels of any one museum. In fact, only two interviewees reported that their museums had successfully infused
evaluative thinking into their organizational cultures, reinforcing Sanders’ (2001) assertion that “we have a long way to go toward developing a culture that will support and invite evaluation in our institutions and organizations” (p. 366).

**Evaluation expectations.** Overall, findings from the first phase of this study indicated that in Canadian art museums most program evaluations were internally driven and not conducted to fulfill funder requirements as may be the case in other accountability contexts (e.g., US, UK). Specifically, while in five museums leadership demand for evaluation was made explicit, the evaluation expectations of Canadian funders varied considerably across the sector. Those funders that required evaluation did not have specific expectations about how art museums should undertake their evaluations or who should be responsible for conducting them—a finding that aligns with those of Hall, Phillips, Meillat, and Pickering’s (2003) research into evaluation in the nonprofit sector in Canada. Further to this, the overwhelming majority of interviewees reported being asked to communicate evaluation results to both funders and their museum’s leadership using output measures alone. This latter finding aligns with other researchers (e.g., Adams, 2012) who have noted that many museum practitioners say that their supervisors, administrators, and/or funders are only interested in output data.

**Responsibility for evaluation.** Akin to findings from research into evaluation practices among voluntary organizations (e.g., Carman, 2007; Hall, Phillips, Meillat, & Pickering, 2003), in the majority of cases, responsibility for conducting evaluations fell to program staff who received varying degrees of support with data collection, analysis, and reporting from colleagues working both within and outside of their departments. Program staff working across the museums sampled dedicated roughly ten percent of their time to evaluation activities—a finding
that exceeds numbers previously reported by previous research into program evaluation in museums (e.g., Parsons, 2004).

**Partnerships with experts in evaluation.** Roughly half of Canadian art museums reported relying on partnerships with various external experts, including university faculty and students, to bolster their own internal capacities for evaluation. In general, such museums were more likely to use more formal data collection, analysis, interpretation, and reporting methods and to conduct both summative studies and outcome evaluations than those museums that did not embrace a partnership approach to evaluation. This finding aligns with observations made by Owen and Visscher (2015), who asserted that conducting evaluation through museum-university partnerships had the potential to be especially beneficial for those museums with limited in-house capacity. It is likewise consistent with research conducted by authors in other contexts (e.g., Cheng & King, 2017; Cousins, Bourgeois, and Associates, 2014; Lobo et al., 2018), who also found that partnerships can help address evaluation knowledge and skills gaps.

**Funding for evaluation.** The overwhelming majority of interviewees reported that they did not receive any financial resources for program evaluation from their institutions. Similarly, only two interviewees reported that their funders either provided monies to support evaluation activities or allowed project funds to be used for evaluation purposes. This finding matches the observations of Hall, Phillips, Meillat, and Pickering (2003), who noted that many funders (i.e., 52%) in the Canadian voluntary sector who demand evaluation do so without recognizing the need to provide financial support.

**Evaluation knowledge and skills.** There is consensus in the evaluation literature that the evaluative knowledge and technical skill level among staff in the voluntary sector related to program evaluation is low (Bozzo, 2002; Carman & Millesen, 2005; Chaytor, MacDonald, &
Melvin, 2002; Hall, Phillips, Meillat, & Pickering, 2003; Juillet, Andrew, Aubry, & Mrenica, 2001). This consistent finding in the literature also turned out to be the case in the Canadian art museum sector. Notwithstanding a few notable exceptions, most interviewees with responsibility for program evaluation lacked the technical skills to conduct what Sanders (2001) described as *disciplined* evaluations or “Planned, systematic, funded evaluations that adhere to professional standards, principles, and codes” (p. 363). Interviewees nonetheless possessed sufficient capacity to conduct informal, mostly output-based, formative evaluations of their programs—a finding that aligns with Adams’ (2012) observation that formative studies are both easier and more realistic for art museum educators to do themselves.

**Use of evaluation findings.** Interviewees with responsibility for evaluating educational programs in Canadian art museums used the results of the program evaluations they conducted for a wide range of purposes. This included using evaluation findings for both problem-solving and decision making purposes (e.g., to restructure camp programs, to make changes to teacher resources, to determine if a program should continue). Interviewees likewise used findings to better understand their target audiences and programming models, to educate their leadership about the value and utility of evaluation, and to identify and share lessons learned about effective practices and programs that might be worthy of application to new settings. Process uses (e.g., learning to think more empirically, acquiring practical evaluation skills) were likewise reported among those interviewees who reported having worked alongside an external evaluator or researcher on a formal study. Overall, the results of this study revealed that Canadian art museums’ capacities to *use* evaluation were higher than their capacities to *do* them. This finding is consistent with research conducted by Cousins, Goh, Elliott, Aubry, and Gilbert (2014), who
found that significantly higher levels of capacity to use evaluation tended to be reported in voluntary organizations than in government organizations.

**Evaluation practice.** Not unlike their counterparts working in the Canadian voluntary sector (Hall, Phillips, Meillat, & Pickering, 2003), interviewees reported engaging in a variety of evaluation activities ranging from informal formative studies to judgment-oriented summative evaluations. A wide range of formal and informal methods were used to collect, analyze, and interpret data, and report evaluation results. Document reviews (e.g., of visitor comments, attendance records, and revenue records) and questionnaires were used most frequently to gather data followed by informal conversations and unstructured observations. Roughly half of the practitioners interviewed reported collecting outcome or impact information. Meaning around the data collected was most often made using either standard qualitative content analysis or simple quantitative analysis techniques. Evaluation findings were disseminated via oral briefings during management meetings as well as through both formal reports and peer-reviewed academic journals. In general, and not surprisingly, practitioners with pre-service or in-service training in evaluation in addition to those who partnered with external experts were more likely to not only use more formal data collection, analysis, and reporting methods, but also to conduct both summative studies and outcome evaluations.

**Satisfaction with evaluation.** All but one of the interviewees I spoke with expressed either being dissatisfied with or interested in improving the quality and quantity of their program evaluations. This finding reinforces a recent observation made by Adams (2012) that “[w]hile the ‘want to’ in the practitioner’s desire to conduct evaluation is much stronger than ever before, the ‘can do’ is a different matter” (p.28).
Engagement with ECB. Not unlike other research that suggests that efforts in the area of ECB in the nonprofit sector remain a work in progress (Cutt & Murray, 2000; Hall, Phillips, Meillat, & Pickering, 2003; Juillet et al., 2001), the findings from this study revealed that ECB is still a relatively unusual occurrence in the Canadian art museum sector. Although a mere three practitioners described having participated in direct ECB activities while in-service, such training had significant long-term impacts on participants. In general, the capacities of those interviewees without formal pre-service training in evaluation who had nevertheless participated in direct ECB were significantly higher than those of their colleagues who had only participated in indirect activities (e.g., conference presentations, networking activities).

Evaluation challenges and needs. A lack of evaluation resources (i.e., time, staff, money) was identified as posing the biggest challenge to interviewees’ abilities to evaluate their educational programs. Other barriers to evaluation that interviewees reported most frequently were both a lack of skills and knowledge in conducting evaluations and institutional demand for and commitment to evaluation. Given these challenges, it is unsurprising that interviewees reported that their biggest need was for more human and financial resources followed by both evaluation training and the sustainable integration of evaluation into their museums’ everyday operations. These findings support some of the challenges and needs that have already been reported in the literature related to evaluation in the voluntary sector (e.g., Blewden, 2010; Bozzo, 2002; Carman & Millesen, 2005; Cutt & Murray, 2000; Hall, Phillips, Meillat, & Pickering, 2003; Juillet et al., 2001; Naccarella et al., 2007).

Concluding comments. Overall, the Phase One results painted a portrait of moderate capacity for program evaluation across the sector with smaller pockets of high capacity. It is important to point out, however, that in line with findings from Bourgeois and Cousins’ (2013)
study, each participating museum’s EC varied from one framework dimension to the next. For instance, despite the fact that in one of the museums participating in the first phase of this study, staff were engaged in direct ECB, such activities had little impact on its overall low evaluation culture. This finding supports Bourgeois and Cousins’ (2013) contention that EC might best be built in organizations using an approach that targets low-capacity areas and focused on developing such areas over time. In their report, Hall, Phillips, Meillat, and Pickering (2003) argued that when conducting research into evaluation in non-government settings, it was imperative for the authors to use a definition of evaluation that accurately captured “the activities of voluntary sector practitioners rather than adhere to the definition found in the more formal field of evaluation research, which relies on social research methodologies” (p. 10). If we agree with Hall, Phillips, Meillat, and Pickering’s (2003) argument, given the fact that all the interviewees in the first phase of this study reported using informal and ongoing approaches to program evaluation, we might go as far as to say, as Carman and Fredericks (2010) have done in their previous research into EC in the nonprofit sector, that when it comes to the program evaluation capacity of Canadian art museums the glass might likewise be “approaching half full” (p. 101).

**Key findings pertaining to research question number three.** The third research question guiding this study was: *How is overall organizational evaluation capacity manifested in those Canadian art museums most active in conducting a wide range of visitor studies, and which factors and conditions, according to practitioners’ self-reported understandings, enabled their capacities to develop to their current levels?* First and foremost, the findings from the second phase of this research provided an in-depth picture of what developed capacity looked like in each of the two art museums selected for case study. Both the MMFA and AGO had high
capacities for research and evaluation as was evidenced through their significant capacities to do evaluation and their ongoing engagement in a various forms of visitor studies inquiry as part of mission fulfillment. While the MMFA had a higher capacity to do and use program evaluation than did the AGO, the latter museum emerged as a leader with respect to exhibition evaluation. In addition to making their EC visible, the cross-case analysis further provided considerable insight into the factors that were at play in shaping these museums’ overall capacities. A summary of these factors in addition to a discussion of the key analytic generalizations or takeaway lessons that I extrapolated from them (Lincoln & Guba, 1985)—articulated as strategies—represent the main contributions of the present section of this chapter.

**Common success factors.** The comparison of the MMFA and the AGO’s individual case portraits identified several common influencing factors that enabled their organizational evaluation capacities to develop to their current levels. The factors that were common to both museums were as follows:

- **Leadership demand.** Long-term leadership interest in and demand for two or more domains of museum visitor studies inquiry were key drivers of visitor research and evaluation at both museums. Specifically, leadership commitment was a principal driving force behind expectations for program evaluation at the MMFA and exhibition evaluation at the AGO.

- **An evaluation-minded culture.** While it did not happen overnight, both the MMFA and the AGO had actively developed cultures that valued research and evaluation. Not only were research and evaluation incorporated at all levels of these institution and across various departments through appropriate structures at both museums, but they were also described as integral to the achievement of certain fundamental institutional values.
Specifically, at the AGO visitor studies were perceived as an avenue to providing an exceptional visitor experience, while at the MMFA research and evaluation were key tools that were used in the service of justice (i.e., to promote social cohesion and individual health and well-being).

- **Resident expertise.** Both museums employed staff with knowledge and skills in at least one of the domains of museum visitor studies inquiry. Specifically, both the MMFA and the AGO hired staff with the requisite knowledge and skills to conduct market research. The AGO further employed highly qualified staff to evaluate the museum’s varied exhibition programs.

- **Direct and indirect ECB.** While at both museums staff were engaged in informal ECB activities (e.g., attending conference presentations on topics related to visitor studies), the AGO further encouraged staff responsible for exhibition evaluation and market research to continue developing their evaluative expertise through participation in direct ECB (i.e., additional certification).

- **Cross-departmental staff support.** Not unlike other Canadian art museums, at both the MMFA and the AGO staff members with responsibility for visitor studies were assisted in their work by colleagues working both within their own and across various other departments. For example, at the MMFA program staff reported being aided by their colleagues working in the museums’ Membership and Customer Services department, while at the AGO such staff noted receiving support from both the museum’s contact centre and its food and beverage team.

- **Committee support.** Staff members with responsibility for conducting visitor studies at both museums were supported by the work of various voluntary committees. Specifically,
at the MMFA program staff were supported by the work of several transdisciplinary advisory committees that were comprised of experts who were leaders in their fields (e.g., the Art and Health Advisory Committee). At the AGO, staff responsible for exhibition evaluation and market research were supported by the work of several internal working groups comprised of staff from across various departments (e.g., the Insights Working Group).

- **Partnerships with external experts.** To varying degrees, both museums undertook studies in partnership with various external experts (e.g., university professors, health institute-based researchers, graduate students, advertising agencies). In fact, the MMFA had institutionalized this strategy to conduct systematic, judgement-oriented, outcomes-based summative studies of their varied programs.

**Key lessons learned.** The factors outlined above provide insight into the specific ways in which art museums can provide the conditions necessary to facilitate the development of their capacities for research and evaluation. Many matched those often found in the evaluation literature (e.g., Cheng & King, 2017; Cousins, Goh, & Clark, 2006; Gullickson, 2010; Katz, Sutherland, & Earl, 2002; Lobo, 2018; Sanders, 2003; Sutherland, 2004). For example, leadership support as a factor to the enculturation of evaluation has been reported in several other studies (e.g., Adams, 2012; Bourgeois & Cousins’, 2013; Cousins et al., 2008; Steele-Inama, 2015; Taylor-Ritzler et al., 2013). The strategy of partnering with external experts on evaluation has also been identified by several scholars as a potential solution for those organizations that want to increase their EC but either lack resident expertise or are unable to hire external evaluators (e.g., Cheng & King, 2017; Cousins, Bourgeois, and Associates, 2014; Lobo et al., 2018). King (2005) likewise found that creating evaluation committees was essential for helping
design, monitor, and reflect on evaluation activities in school districts. This research therefore aligns with and supports the role that leadership as well as other well documented factors such as culture, expertise, ECB, and human resources play in positively influencing EC in organizations while also adding to our understanding of how such factors are manifested in the context of art museums.

Conversely, this research also suggests that external pressure and financial resources are factors that might not impact developed capacity in the Canadian art museum sector to the extent that they have been reported to do so in other sectors (i.e., Canadian government) or accountability contexts (e.g., U.S. art museums). Based on my examination of the above influencing factors, I was able to extrapolate several tangible strategies that art museums could employ to bolster their capacities around evaluation. These included (a) employing qualified staff to undertake studies internally, (b) partnering with external experts in research and evaluation, (c) forming evaluation committees, and (d) engaging staff in direct ECB. While not meant to be mutually exclusive or exhaustive, it is my contention that these strategies are worthy of adaptation and application to other art museums and, depending how closely their own situations match those of the case studies, may also be transferrable to other contexts such as arts-based nonprofit organizations and other museum types.

**Implications of Key Findings**

The findings of this inter-disciplinary research have implications for both art museum education and program evaluation theory, practice, and research. These implications are summarized and discussed in the paragraphs below.

**Implications for the framework.** This study provided the evidence necessary to complete the initial *Conceptual Framework of Evaluation Capacity in Art Museums* reviewed
earlier in this section. Specifically, in light of the findings pertaining to research question number one, I made several adjustments, reflected in Figure 5 below, to the initial conceptual framework in an effort both to better outline the key dimensions of EC that were actually observed in Canadian art museums and to reflect the relationships that emerged among them. After presenting each adjustment in turn, I discuss the contribution the final framework makes to the theoretical literature base related to the conceptualization of EC. The adjustments were as follows:

- *Context* and *culture* are now conceptualized as a single dimension. Hence, the first category in my revised framework consists of only three dimensions, conceptualized as interrelated influencing factors—*context and culture, demand for evaluation,* and *resources and supports.*
- An arrow indicates a reciprocal relationship between those dimensions conceptualized as influencing factors and those conceptualized as capacity indicators to allow for the impact that continuous engagement with evaluation, for instance, might have on a museum’s evaluation culture.
- The original dimension *capacity to do and use program evaluation* was parsed out into two separate dimensions for clarity—*capacity to do evaluation* and *capacity to use evaluation.*
- The revised dimensions *capacity to do evaluation, capacity to use evaluation,* and *evaluation activity* are no longer depicted as nested constructs. This is not to suggest that program evaluation in museums does not take place within the larger body of research knowledge production known as visitor studies, but rather to express these constructs more generically, at a higher categorization level. The change was made to enable
museum practitioners and researchers alike to use the framework regardless of the domain of visitor studies inquiry they might choose to investigate in comparison to others (i.e., scholarly research, market research, exhibition evaluation, program evaluation).

- ECB is no longer conceptualized as a dimension unto itself. Rather, it is embedded within the knowledge and skills sub-dimension falling under the capacity to do evaluation dimension and as a sub-dimension of the resources and supports dimension.
- Being too numerous to list, all references to sub-dimensions have been removed from the framework graphic itself. The sub-dimensions that comprise each individual dimension can be articulated in future framework descriptions.
- Soft skills, such as interpersonal skills and knowledge of audiences, programs, and informal learning have been added under the capacity to do evaluation dimension to reflect the assets practitioners reported bringing to program evaluation in Phase One.

Figure 5. The conceptual framework of evaluation capacity in art museums (revised)

Several scholars have called for more research that advances the evaluation field’s theoretical understanding of EC across a variety of sectors and organizational types (e.g.,
Cousins & Bourgeois, 2014; Cousins, Goh, Elliott, & Bourgeois, 2014; Taylor-Ritzler et al., 2013). The revised framework now clearly demarcates what EC looks like in the context of Canadian art museums. It therefore represents an important and original contribution to evaluation theory because no empirical data were previously available on the actual dimensions of EC that would manifest in these professional settings. While the framework is well suited to both understanding and describing evaluation practice and capacity in the art museum sector, with minor adjustments, it also has the potential to cater to the needs and circumstances of other types of museums (e.g., science and natural history museums), thus promising to be of value to the broader museum community. Given the fact that some of the findings of this study, discussed earlier in this chapter, parallel those of previous research into evaluation in the nonprofit sector, the Conceptual Framework of Evaluation Capacity in Art Museums might likewise be useful to Canadian arts and culture organizations as well as education-sector nonprofits whose situations are likely to echo those of many Canadian art museums. The implications of the framework for future research are discussed towards the end of this chapter.

It is important to note, as I have likewise done in Chapter Two, that the Conceptual Framework of Evaluation Capacity in Art Museums was necessarily developed based on a literature review of EC frameworks and studies published up to and including 2016 in order to guide instrument development prior to both seeking ethics clearance and recruiting Phase One participants beginning in August of 2017. Despite my serious attention to the integration and synthesis of the available EC literature at the time, as is the case with the normal course of scholarly output, I am aware that other scholars have published new frameworks and studies across a range of different contexts (e.g., Cheng & King, 2017; Fierro & Christie, 2017; Lindeman et al., 2018; Schwarzman, 2019) that may or may not have implications for the
dimensions and sub-dimensions that I include in the revised *Conceptual Framework of Evaluation Capacity in Art Museums*.

For example, in their 2017 study Cheng and King identified both the major dimensions of EC and feasible approached to ECB in Taiwanese elementary and junior high schools. While the results of the authors’ study support several of the dimensions and approaches already identified in the Western literature (e.g., culture, financial and human resources), because the final version of their EC/ECB framework was identified as suitable for Taiwanese schools, I do not expect it to alter the conclusions I have made in the present study. This said, future work, could consider the complementarity of emerging EC frameworks as they relate across contexts.

**Implications for practice.** On the basis of the findings of this study, I have drawn out a number of key lessons that I here frame as recommendations for action, outlining what could be done to strengthen the evaluation capacity of art museums in Canada. These recommendations include training leaders, training art museum educators, developing an evaluation website or clearinghouse, forging museum-expert partnerships in research and evaluation, creating museum networks around evaluation, and increasing both demand and funding for evaluation across the sector.

**Training leaders.** The finding of the second phase of this study illustrated just how critical an art museum’s leadership was to the enculturation and use of evaluation in this sector. Unfortunately, the phase one findings also revealed that most Canadian art museum Directors had yet to realize the value and utility of evaluation. This finding is perhaps unsurprising given the fact that the overwhelming majority of art museum Directors come from curatorial backgrounds where they would not have received any formal training in evaluation as part of their professional preparation.
In addition to a dearth of pre-service training, such leaders also have few evaluation-related in-service training opportunities available to them. Specifically, professional associations such as the Canadian Evaluation Society (CES) do little to reach leaders in museums, while others such as the Canadian Art Museum Directors Association (CAMDO) and the Canadian Museum Association (CMA) rarely address evaluation at their meetings. As a result, as the case of the AGO showed, negative attitudes towards evaluation can still proliferate among museum professionals who come from such backgrounds.

Sanders (2001) has argued that the development of the arts is an example of a program that “should embrace disciplined evaluation” (p. 364). Therefore, not only could Canadian universities better prepare future curators for leadership roles in museums by providing a course in evaluation in their faculties of art history, fine arts, and museum studies but also professional associations (e.g., CAMDO, CES, CMA) could ally themselves in educating museum leaders about the importance of evaluation. Moreover, ECB developers could be employed to work directly with museum Directors to change the way they think about evaluation and illustrate its potential and value. Directors the likes of the MMFA’s Nathalie Bondil, who have successfully built an evaluation culture in their institutions, could also play a key role in helping other leaders recognize how evaluative inquiry could aid them in realizing their own museums’ missions.

**Training art museum educators.** The findings of this study revealed that responsibility for evaluating educational and public programs in Canadian art museums primarily fell to program staff who had received little to no training in evaluation and, hence, struggled to demonstrate the value of programs beyond outputs. While this finding aligns with previous research into the qualifications of art museum educators (e.g., Chen-Cooper, 2007; Ebitz, 2005), it also points to the importance of building such practitioners’ knowledge and skills around
evaluation through a combination of both pre-service education and in-service training.

Hein (1994) has argued that evaluation work not only helps to strengthen, but should also be a part of any museum education program. Program evaluation should therefore be a required course of study in the preparation of all Canadian art museum educators. Institutions of higher education such as Concordia University, which is in the process of inaugurating a new graduate certificate program in museum education, could lead the way by equipping graduate students with the full range of skills necessary to succeed in their future roles as museum educators—skills that, as this research has shown, clearly include program evaluation and other related forms of visitor research. Universities could likewise help to curtail the low EC rates that pervade amongst Canadian art museum educators by offering one-off, stand-alone courses not attached to degrees for those practitioners interested in developing evaluative skills, but who are unable to leave their jobs to do so.

Professional associations such as the Visitor Studies Association, the American Evaluation Association—which has an Arts, Culture, and Audiences topical interest group (TIG)—and the Canadian Evaluation Association could do a better job of targeting museum educators with their professional development offerings. Other associations such as the Canadian Society for Education through Art and the Canadian Art Gallery Educators network, meanwhile, could begin offering conference workshops on evaluation.

Despite still being rare occurrences within the Canadian art museum sector, the results of this study also point to the fact that evaluation approaches that are characterized by the engagement of participants in evaluation processes (e.g., participatory, collaborative, and empowerment evaluation) and on-site face-to-face training workshops with evaluation experts were effective ways of building art museum educators’ hard skills in evaluation. More Canadian
art museums might therefore consider both involving program staff in this type of in-service training and using approaches to evaluation where the evaluation process is used to intentionally build capacity as two possible paths towards the ultimate goal of developed capacity.

**Development of a central portal to evaluation projects, research, and resources.** Further to expressing a need and desire for more pre-service and in-service training in evaluation, the art museum educators interviewed as part of this study described needing access to evaluation reports from other art museums in addition to examples of evaluation techniques and instruments to guide them in their evaluative work. Many also felt frustrated because they could not readily access such information. This finding both aligns with Adams’ (2012) contention that “getting more report findings out to practitioners is critical” (p. 33) and points to the need for a central portal, website, or clearinghouse that could serve as a vehicle for (a) sharing evaluation reports, practices, and experiences, (b) locating an evaluator, researcher, ECB developer, or other external expert, and (c) accessing evaluation resources (e.g., toolkits, handbooks, databases, associations, example instruments). A similar need has likewise been identified by practitioners working in the Canadian voluntary sector (Hall, Phillips, Meillat, & Pickering, 2003). In Canada, the development of such a portal may be of interest to the Canadian Museums Association. More broadly, it may also be of relevance to the membership of the International Council of Museums (ICOM).

**Museum-expert partnerships in research and evaluation.** Owen and Visscher (2015) argued that conducting visitor research and evaluation studies in partnership with students enrolled in graduate programs (e.g., museum studies, education, market research) at local universities could contribute towards enhancing a museum’s internal evaluative efforts. The findings of this research further suggested that, in addition to this approach, EC in the Canadian
art museum sector could be increased through the establishment of evaluative partnerships with a wide range of other external experts. These experts included individuals who were affiliated with universities, hospitals, health agencies, nonprofit organizations, and both evaluation consulting and media firms (e.g., academic faculty, researchers, professional evaluators, marketing consultants, and health care professionals such as nurses, physicians, and therapists).

While even one-off partnerships proved to be beneficial (e.g., practitioners were enabled to conduct rigorous evaluations into programs and honed their skills in social science and evaluation methods), the case of the MMFA demonstrated that, when adopted broadly, a partnership approach was a compelling strategy towards developing a museum’s capacity for evaluation. Specifically, the approach permitted the MMFA to (a) circumvent limitations posed by a lack of dedicated evaluation funds, (b) carry out both more and higher quality studies than it could have otherwise conducted in an internal capacity, (c) capitalize on a wide range of academic and other expertise, (d), inspire greater confidence in its research and evaluation findings, and (e) strengthen its ability to disseminate its collective impact. Hence, the strategy of partnering with external experts in research and evaluation can serve as a powerful model for other Canadian art museums interested in increasing their capacities around evaluation.

**Evaluation networks.** The findings of this study indicated that opportunities exist for Canadian art museums to create networks around evaluation, a strategy that has been quite successful in the museum field in the United States (e.g., Steele-Inama, 2015). Existing networks such as the collaboration of twelve museums in Balboa Park, San Diego or the Denver Evaluation Network (DEN, 2019) could therefore serve as models for those Canadian art museum practitioners who expressed an interest in starting and sustaining a similar network on this side of the border. These types of efforts could focus on building the evaluative capacities of
art museum practitioners through a variety of multidisciplinary activities. For instance, museum networks could pool their resources to fund evaluation workshops or other related direct ECB activities for their members. Such workshops might focus on helping practitioners demonstrate measurable outcomes of programs and could be facilitated either by recognized evaluation experts or those individuals within the network who already possess evaluative knowledge and skills. Individuals with evaluative expertise could also be enlisted to share effective evidence-based practices with as well as provide technical assistance and mentoring to their less experienced counterparts. Further to this, network members could assist one another in finding relevant external experts to partner with on research and evaluation studies as well as collaborate to conduct pan-institutional studies together.

**More demand and concomitant funding for evaluation.** The findings of this research revealed a desire among Canadian art museum practitioners for both increased institutional and funder demand for evaluation. Specifically, while some practitioners expressed wishing to see their museum’s leadership request and use evaluation data more routinely in their governance, others believed that a consistent requirement from funders would be the best way to sustain evaluation across the sector. Unlike in the Canadian government sector, there is no policy in the museum sector that sets out the fundamental requirement for institutional accountability for performance information and evaluation. The Museum Act (GOC, 1990), which governs those museums that operate as federal Crown corporation, could therefore be amended to serve as such a policy and lead to greater demand for evaluation across the museum sector.

The findings of this research also point to the fact that with respect to funding for evaluation, there was much room for improvement in Canada’s art museums. Not only were institutional expectations for evaluation not accompanied by financial support, but also, of those
funders that demanded evaluation, a mere two provided funds to be used for such purposes. Canadian art museums should therefore lobby their funders to either include funding for evaluation in any grants they provide or allow them to direct a portion of a total grant value towards evaluation. They should also start including requests for evaluation funds in their grant applications. Those museums with a strong research agenda (e.g., the MMFA), could further consider lobbying Canadian federal and provincial research funding agencies (e.g., the Fonds de recherche du Québec, the Social Science and Humanities Research Council of Canada) to revisit their applicant eligibility criteria and make it easier for such museums to apply for research grants.

**Implications for research.** Both the constraints of this study and its findings raised a number of different considerations for future research. First, because this study focused on exploring capacity for evaluation in Canada’s largest public and government-owned art museums, it both ignored art museums in other national and accountability contexts as well as other museum types. Since museums are not a homogenous group and Canadian art museums are unlikely to be representative of art museums as a whole, additional research is still needed that examines what EC looks like both within art museums outside of Canada (e.g., US, UK) and across a range of different types of institutions (e.g., art, science, and heritage).

For example, the revised *Conceptual Framework of Evaluation Capacity in Art Museums* could be applied to the examination of similarities and differences between art museums’ capacities for evaluation in two distinct national contexts: Canada and the US. Similarly, the framework might be both adapted and applied to explore how EC is manifested within Canada’s six national museums (i.e., the Canadian Human Rights Museum, the Canadian Museum of History, the Canadian Museum of Immigration at Pier 21, the Canadian Museum of Nature, the
National Gallery of Canada, and the Canada Science and Technology Museum). Such comparative research will be vital to better understanding the validity of the framework and the influencing factors that could support greater EC across contexts and in both art and non-art museums.

As this research has revealed, art museums are dynamic and ever-changing organizations whose capacities for evaluation are likely to evolve over time. Because of the cross-sectional nature of this study, another potential avenue for future research would therefore be to conduct a follow-up study in perhaps ten years’ time to track any changes in Canadian art museums’ EC and understand the factors and conditions that might serve to bring such changes about. Longitudinal research designs, as Weiss (1981) has previously argued, are ideally suited to the study of complex, dynamic social processes. Such a design could therefore be invaluable to advancing understandings of the complex phenomenon of organizational capacity for evaluation in the context of art museums.

This research also points to several key promising practices in art museum education evaluation that are worthy of continued exploration. Partnering, for example, emerged as an organizational strategy with strong potential for developing an art museum’s internal capacity for program evaluation. Future research might therefore explore the permutations in partnership models that exist across sectors resulting in a toolkit that might allow art museums to follow a step-by-step process for building a sustained evaluation partnership with two or more organizations representing both the university and other sectors. Research in this area might also examine the extent to which those museum practitioners who partner with external experts benefit from engaging in evaluation process (e.g., by and acquiring practical evaluation skills).
Because a two-phase multiple method qualitative research design was used to address the purposes of this research, I was only able to extract analytic generalizations or key takeaway lessons learned (Lincoln & Guba, 1985) from the findings. Continuing research into EC in art museums could therefore use quantitative research methods (e.g., Confirmatory Factor Analysis, Structural Equation Modeling) in order to (a) move beyond analytic to statistical generalizability (Nielsen, Lemire, & Skov, 2011) and (b) further explore the relationship between the dimensions conceptualized as influencing factors and capacity indicators in the Conceptual Framework of Evaluation Capacity in Art Museums. The results from such analyses would further provide statistical evidence for the use of the framework to measure organizational capacity for evaluation in the art museum sector.

Further to the research directions suggested above, additional research might serve to answer some of the following questions that were raised throughout this study:

1. What do Canadian funders’ evaluation expectations look like across the country? Why do so few Canadian funders require the museums they fund to conduct evaluations, and why do the majority of them seem to only be interested in output data? Is it, as Hall, Phillips, Meillat, and Pickering (2003) have suggested, that funders themselves lack the capacity to review and act on evaluation information? Why do those funders that demand evaluation not recognize the need to provide art museums with financial support for this?

2. What does the pre-service preparation of art museum educators look like across Canada? Why does program evaluation have such limited visibility in such preparation? How might the inclusion of an evaluation course in such preparation translate to art museum educators’ in-service practice?
3. What is the link between disciplined evaluation in art museums and both program success and organizational performance?

4. Why, despite museum education’s rise in stature over the last few decades, is education not central to the missions of more Canadian art museums?

Conclusion

As the first pan-Canadian study of its kind, this research makes an important and original contribution to art museum education and both visitor studies and program evaluation theory, research, and practice. First, in providing empirical support for the dimensions included in my conceptual framework, this research contributes to the theoretical literature base related to the conceptualization of organizational evaluation capacity. Second, this research represents a critical step towards addressing a pressing research gap by advancing knowledge on both program evaluation and other domains of visitor studies inquiry (i.e., market research, exhibition evaluation) in the context of Canadian art museums. As such, it adds to more general understandings of both assessment practices in the field of art education and audience research and evaluation practices in the field of museum studies. Lastly, this research offers art museums assistance in the form of examples of institutional successes and recommendations for practice that outline what could be done to strengthen the evaluation capacity of art museums in Canada. Such assistance is likely to be useful not only to the growing number of art museums seeking to integrate evaluation into their organizational cultures, but potentially also to several other sectors and organizational types (e.g., non-art museums and both education-sector and arts-based nonprofit organizations).
References


https://doi.org/10.1177/1098214013477235


https://doi.org/10.1016/j.evalprogplan.2015.01.004


Cousins, J. B., & Bourgeois, I. (Eds.) (2014). Organizational capacity to do and use evaluation [Special Issue]. *New Directions for Evaluation, 141*.


https://doi.org/10.1080/00393541.2005.11651787

https://doi.org/10.1111/j.2151-6952.2000.tb01155.x


https://www.jstor.org/stable/1321039

91.


Western Michigan University, Kalamazoo, MI.


https://doi.org/10.2307/3193592


Evaluation and Program Planning, 30(3), 231–236.
https://doi.org/10.1016/j.evalprogplan.2007.05.001


https://doi.org/10.1177/1098214010396075


Reid, N. (2012). The contemporary identities of art museum educators as told through their life histories (Unpublished doctoral dissertation). Concordia University, Montreal, QC.


https://doi.org/10.1080/1028663032000089868


Evaluation, 23(3), 253–259. https://doi.org/10.1016/S1098-2140(02)00199-6


Appendix A
Letter of Ethics Clearance from Queen's University

August 09, 2017

Ms. Agnieszka Chalas
Ph.D. Candidate
Queen's University
Kingston, ON, K7L 3N6

GREB Ref #: GEDUC-864-17; TRAQ # 6021509
Title: "GEDUC-864-17 Painting a Portrait of Organizational Evaluation Capacity in the Canadian Art Museum Sector"

Dear Ms. Chalas:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GEDUC-864-17 Painting a Portrait of Organizational Evaluation Capacity in the Canadian Art Museum Sector" for ethical compliance with the Tri-Council Guidelines (TCPS 2 (2014)) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (405.001), your project has been cleared for one year. You are reminded of your obligation to submit an annual renewal form prior to the annual renewal due date (access this form at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Annual Renewal/Closure Form for Cleared Studies"). Please note that when your research project is completed, you need to submit an Annual Renewal/Closure Form in Romeo/traq indicating that the project is 'completed' so that the file can be closed. This should be submitted at the time of completion; there is no need to wait until the annual renewal due date.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period (access this form at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Adverse Event Form"). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example, you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To submit an amendment form, access the application by at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Request for the Amendment of Approved Studies". Once submitted, these changes will automatically be sent to the Ethics Coordinator, Ms. Gail Irving, at the Office of Research Services for further review and clearance by the GREB or GREB Chair.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Sincerely,

Joan Stevenson, Ph.D.
Interim Chair
General Research Ethics Board

c: Dr. Christopher DeLuca, Supervisor
   Dr. Richard Reeve, Chair, Unit REB
   Ms. Erin Rennie, Dept. Admin.
Appendix B
Sample Email Scripts

For Phase One (heads of education, internal evaluators)
My name is Agnieszka Chalas and I am undertaking a research project for my PhD at Queen’s University Faculty of Education. I would like to invite you to participate in a pan-Canadian research study about program evaluation in art museums. The overarching purpose of this study is to develop a deeper understanding of Canadian art museums’ evaluation practices and capacities than is currently the case. As such, understanding your perspective as a museum professional whose job responsibilities include program evaluation (internal evaluator, museum educator, head of education/public programming* select and delete as appropriate) is important to this research. You will be invited to undertake one interview lasting a total of no more than 60 minutes. Interview questions will focus on your experiences with evaluating educational programs in addition to your evaluation challenges and resource needs. Your identity will not be disclosed in any presentation of results and your confidentiality will be protected to the best of my ability. Your participation in the study is voluntary. Should you agree to participate, I will provide you with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in the study. Please contact me if you need more information about this study.

For Phase Two (known museum contact persons i.e., key informants interviewed during Phase One)
Thank you again for allowing me to interview you as part of my doctoral study, Painting a Portrait of Program Evaluation in Canadian Art Museums. Based on the results of the interview study you recently participated in, your museum was identified as one of two art museums in the country most active in evaluation and, as a result, was selected for in-depth case study. The overarching purpose of the case study research is to develop a better understanding of what ‘best practices’ evaluation looks like at these two art museums so that such practices can be pinpointed and shared across the sector. I would like to invite you to assist me in identifying additional staff members who might be good sources of information about program evaluation or visitor studies at your museum that I may approach about participating in this second phase of my research. Despite the inclusion of your museum in this study, your decision to refer colleagues is voluntary. Should you choose to recommend possible staff members by responding to this email, the individuals you name will be provided with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in this research. Please contact me if you need more information about this study.

For Phase Two (Executive Directors)
My name is Agnieszka Chalas and I am undertaking a research project for my PhD at Queen’s University’s Faculty of Education. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in the country most active in evaluation and, as a result, was selected for in-depth case study. The overarching purpose of the case study research is to develop a better understanding of what ‘best practice’ program evaluation looks like at these two art museums so that such practices can be pinpointed and shared across the sector. The purpose of this email is to seek your permission to include the _______ as one of the
two museums in my case study and to invite you to participate in a 60 minute in-person interview about your museum. During the interview, I also hope to collect any relevant documents that may add to my understandings of your museum. Your decision to participate and include the ______ in this study is voluntary. Should you agree to participate, I will provide you with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in the case study as well as your options around disclosure including the option not to be identified. Please contact me if you need more information about this study.

**For Phase Two (prospective participants)**

My name is Agnieszka Chalas and I am undertaking a research project for my PhD at Queen’s University Faculty of Education. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in the country most active in evaluation and, as a result, was selected for in-depth case study. The overarching purpose of the case study research is to develop a better understanding of what ‘best practice’ program evaluation looks like at these two art museums so that such practices can be pinpointed and shared across the sector. I got your name and contact information through ______ who thought you might be a good source of additional information about (program evaluation/visitor studies) at your museum. Your Executive Director has agreed for the ______ to be one of the two museums in my case study. I would therefore like to invite you to participate in a 30-45 minute in-person interview about your experiences with evaluating (educational/exhibition) programs at the ______. During the interview, I also hope to collect any relevant documents that may add to my understanding of evaluation at your museum. Despite the inclusion of your museum in this study and your referral to me by one of your colleagues, your participation in the study is voluntary. Should you agree to participate, I will provide you with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in the case study as well as your options around disclosure including the option not to be identified. Please contact me if you need more information about this study.
Appendix C
Example Telephone Scripts

Phase One: Interview Study

Hello. This is Agnieszka Chalas from Queen’s University. I got your number through ________. I am conducting a pan-Canadian research study on program evaluation in art museums to increase the field’s understandings of these museums’ evaluation practices and capacity for evaluation. You were purposefully selected for participation in this study because, as a museum professional whose job responsibilities include evaluating educational programs in museums, your perspective is important to this research. The interview will take up to 60 minutes. Your participation in the study is voluntary. Are you interested in participating in this study either now or later when you have more time?

If NO = Is there someone else that you would recommend I speak to at your institution instead?  
If No = Thanks and have a good day.  
If YES = Do you have time now, or would you like me to call later?  
If NO – When would be a convenient time to call back?  
If YES – Great. Because this is a research study, I need to take a moment to tell you about the project and your rights as a volunteer.

I wish to gather information about your experiences evaluating museum education programs at your institution. I have 25 questions that will take up to 60 minutes depending on your time and interest. It is important that you first understand that you are not required to answer any questions that you may not wish to answer and that you are able to withdraw from this interview whenever you wish to withdraw with no penalty. There are no known risks to your involvement in answering these questions, your identity will not be disclosed in any presentation of results and your confidentiality will be protected to the best of my ability. Do you understand these rights or want any of them explained further?

YES or NO

I would now like to turn on a tape recorder so that I do not miss any of your comments and I can record your willingness to continue. Once I remove the script from the tapes, I will store our conversation in a protected computer file for five years and erase the tapes. Are these conditions okay for you?

YES or NO (adjust accordingly)  
Turn on/off tape recorder.

Thank you. I am currently talking to [insert name] on [insert date]. Do I have your consent to continue with the interview? I would now like to ask you a series of 25 questions. (ASK QUESTIONS)

Those are all the questions I have for today. It is important that I leave you with contact information should you have any concerns about this interview. Can I leave you with some contact information (phone number or email)? If you wish to speak to me, my name is Agnieszka
Chalas and I can be reached at a.chalas@queensu.ca or at 613-876-5427. If you have any ethics concerns you can contact the General Research Ethics Board (GREB) at 1-844-535-2988 (Toll free in North America) or chair.GREB@queensu.ca. Do you have any questions? Thanks so much for taking the time to participate in this interview.

Phase Two: Case Study (known museum contact persons i.e., key informants interviewed during Phase One)

Thank you again for allowing me to interview you as part of my doctoral study, *Painting a Portrait of Program Evaluation in Canadian Art Museums*. Based on the results of the interview study you recently participated in, your museum was identified as one of two art museums in the country most active in evaluation and, as a result, was selected for in-depth case study. The overarching purpose of the case study research is to develop a better understanding of what ‘best practices’ evaluation looks like at these two art museums so that such practices can be pinpointed and shared across the sector. Your Executive Director has agreed for the _______ to be one of the two museums in my case study. I am therefore hoping that you could assist me in identifying additional staff members who might be good sources of information about program evaluation or visitor studies at your art museum that I may approach about participating in this research. Despite the inclusion of your museum in this study, your decision to refer colleagues is voluntary. Should you choose to recommend possible staff members, the individuals you name will be provided with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in the study and their options around disclosure including the option not to be identified. Are you willing to recommend any additional staff members?

If NO = Thanks and have a good day. (Proceed with identifying participants in a different way i.e., through the Executive Director)
If YES – Great thank you. I appreciate your help in this regard.

Phase Two: Case Study (Executive Directors)

Hello this is Agnieszka Chalas from Queen’s University. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in the country most active in evaluation and, as a result, was selected for in-depth case study. The overarching purpose of the case study research is to develop a better understanding of what ‘best practice’ program evaluation looks like at these two art museums so that such practices can be pinpointed and shared across the sector. The purpose of this phone call is to seek your permission to include the _______ as one of the two museums in my case study and to invite you to participate in a 30-45 minute in-person interview about your museum. During the interview, I also hope to collect any relevant documents that may add to my understanding of your museum. Your decision to participate and include the _______ in this study is voluntary. Should you agree to participate, I will provide you with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in the case study and your options around disclosure including the option not to be identified. Would you be interested in participating in this study?
Phase Two: Case Study (prospective participants)

Hello this is Agnieszka Chalas from Queen’s University. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in the country most active in evaluation and, as a result, was selected for in-depth case study. The overarching purpose of the case study research is to develop a better understanding of what ‘best practice’ program evaluation looks like at these two art museums so that such practices can be pinpointed and shared across the sector. I got your name through ______ who thought you might be a good source of additional information about (program evaluation/visitor studies) at your museum. Your Executive Director has agreed for the ______ to be one of the three museums in my case study. I would therefore like to invite you to participate in a 30-45 minute in-person interview about your experiences with evaluating (educational/exhibition) programs at the ______. During the interview, I also hope to collect any relevant documents that may add to my understanding of evaluation at your museum. Despite the inclusion of your museum in this study and your referral to me by one of your colleagues, your participation in the study is voluntary. Should you agree to participate, I will provide you with a Combined Letter of Information and Consent Form further detailing the purpose of and expectations for participation in the case study as well as your options around disclosure including the option not to be identified. Would you be interested in participating in this study?

If NO = Is there someone else that you would recommend that I speak to at your institution instead?
If NO = Thanks and have a good day.
If YES – Great. When might be a good time for me to interview you in person?
Appendix D
Phase One Combined Letter of Information and Consent Form

Study Title: Painting a Portrait of Program Evaluation in Canadian Art Museums
Name of Student Researcher: Agnieszka Chalas, Queen’s University
Name of Supervisor: Chris DeLuca, Queen’s University

I am Agnieszka Chalas, a doctoral candidate in the Faculty of Education at Queen’s University working under the supervision of Dr. Chris DeLuca. I would like to invite you to participate in a research study entitled *Painting a Portrait of Program Evaluation in Canadian Art Museums.* This study is funded by the Social Sciences and Humanities Research Council of Canada and seeks to understand how program evaluation and visitor research is being conducted across Canada’s largest publicly owned art museums with particular emphasis placed on how such museums are evaluating their *educational* programs. There are no known risks to your participation. While there are no direct benefits to you as a participant, this study will establish an initial knowledge base on program evaluation in Canadian art museums and will be used to share knowledge aimed at improving evaluation practice and increasing capacity for evaluation across the sector.

If you agree to take part, I will interview you for up to 60 minutes over the telephone. With your permission, the interview will be audio-recorded and later transcribed. Your participation in the study is voluntary—you don’t have to answer any questions you don’t want to, can stop participating in the interview at any time, and are free to withdraw from the study up until three months following the interview by contacting me at a.chalas@queensu.ca. If you choose to withdraw, you may request removal of all or part of your data from the study.

The findings from this study may be published in academic journals and presented at professional conferences but any such publications/presentations will protect your confidentiality to the extent possible. For example, your name will not be used in the data or published work. Additionally, your data will be retained for a minimum of five years on a password protected computer in accordance with the Faculty of Education’s policy and no one other than me will have access to any of the data.

If you have any questions about this study, please contact me, Agnieszka Chalas, at a.chalas@queensu.ca / 613-876-5427, or my supervisor Dr. Chris DeLuca at cdeluca@queensu.ca / 613-533-6000 ext. 77675. Any ethical concerns about the study may be directed to the General Research Ethics Board (GREB) at 1-844-535-2988 (Toll free in North America) or chair.GREB@queensu.ca.

Sincerely,

Agnieszka Chalas, Ph.D. Candidate, Faculty of Education, Queen’s University
Keep one copy of the Letter of Information and Consent Form for your records and return one copy to the researcher, Agnieszka Chalas, a.chalas@queensu.ca.

The above Letter of Information provides you with the necessary details to help you make an informed choice about participation. All your questions should be answered to your satisfaction before you decide whether or not to participate in this research study.

By signing below, I am verifying that: I have read the Letter of Information, agree to participate in this study, and all of my questions have been answered.

Name: ________________________________

Signature: ________________________________

Date: ________________________________
Appendix E
Phase One Interview Protocol

Introduction

TO BE READ AND RECORDED
Thank you again for agreeing to speak with me. As you are aware, I am conducting this interview today as part of the research project, Painting a Portrait of Program Evaluation in Canadian Art Museums. I have 25 questions that I will be asking you during this interview which I will be recording and then transcribing verbatim. Are you in agreement with the collection and use of this data to further the research project?

Interview Questions

Context/Culture

I want to begin by asking you a couple of questions about your museum and its culture.

1. To what extent do you feel educational endeavours are valued and/or prioritized at your museum. How would you describe the status of the education department within the overall structure of the museum in which you work?
2. Does your museum have an individual or department in place whose sole responsibility it is to conduct evaluation? If not, who is responsible for evaluating educational programs at your museum?
   Probe
   (a) What percentage of your/their time is dedicated to evaluation?
   (b) When evaluating programs, are there specific standards of practice that you/they subscribe to in your work such as the Visitor Studies Association’s Evaluator Competencies?
   (c) Are the number of people with this responsibility or the percentage of time they spend on evaluation sufficient?
   (d) To what degree do you rely on external evaluators?
3. How important do you think it is to evaluate educational programs at your museum and to what extent is this a priority for you?
4. In this next question I’m trying to ascertain the degree to which there is broad-based interest in evaluation at your museum. How would you describe the evaluation culture at your museum?
   Probe
   (a) How would you describe your colleagues’ attitudes towards evaluation?
   (b) How committed would you say the museum leadership (i.e., ED or Board of Directors) is to evaluation?
   (c) Does your museum have an institution-wide evaluation policy or plan or agenda in place?
   (d) Are you aware of any external policies that govern program evaluation in your museum?
Demand for Evaluation/ Resources and Supports

The next set of questions focus on the demands placed on you for evaluation and the supports available.

1. Does your museum leadership require you to evaluate your educational programs? If so, what kind of evaluation support do you receive (financial or otherwise) from your administration?
   
   Probe
   
   (a) Does your leadership provide you with resources (i.e., time, staff, an evaluation budget, training) for evaluation?

2. How are your educational programs funded and who are your biggest funders?

3. Do your funders require you to evaluate your educational programs? Which percentage of your funders require you to do so? How many would that be?
   
   Probe (if YES)
   
   (a) Do those that require evaluation provide dedicated monies for evaluation or allow project funds to be used for evaluation purposes? What percentage of the grant total is dedicated to supporting evaluation activities?
   (b) Do funders require a formal evaluation plan from the program’s inception?
   (c) Do funders specify the methods they want you to use or who should be responsible for conducting the evaluation?
   (d) What kind of evaluation information do funders expect (e.g., evaluations that report on the outcomes of the programs vs. those that report on outputs)? To what degree is this information that you would have collected regardless?
   (e) To what degree do funders provide other supports (i.e., access to evaluation training or funding for training, advice on evaluation, technical assistance)?

4. What kind of software do you use to compile, analyze, and manage evaluation information?

5. To what degree has meeting funders’ or your administration’s demands for evaluation challenging for you? What barriers do you face to conducting evaluation? Explain.

6. What kinds of supports, whether resource or financial, do you require from funders and your administration to effectively evaluate your programs?

Capacity to Do and Use Evaluation

The following questions ask you to comment on your evaluation capacity and how you use evaluation results at your museum.

1. Describe for me your background.

2. Can you tell me about any specific training in evaluation you may have received?

3. Can you tell me about your previous practical evaluation experience (i.e., number of evaluations, years of experience, working alongside an evaluator)?

4. What do you think is the greatest strength you bring to the evaluation of museum education programs?

5. How comfortable or experienced do you feel with the following:
   
   (a) Framing research and evaluation questions, planning evaluation studies, addressing ethical issues, and designing instruments?
   (a) Collecting, analyzing, and interpreting data as well as reporting evaluation findings?

6. How do you use the evaluation information you collect? For what purposes?
Probe

(a) How are evaluation findings shared across the museum and/or disseminated beyond the museum?
(b) Who reviews program evaluations at the museum (i.e., Board of Directors)?
(c) To what degree do evaluation findings feed into strategic planning?
(d) To what degree are evaluation recommendations acted upon?
(e) To what degree are you able to use evaluation data collected to make decisions about programs?
(f) What did you learn from your last evaluation?
(g) To what degree do you think staff attitudes have changed or their evaluation skills have increased as a result of participating in an evaluation?

7. What would you say are the biggest challenges to using evaluation findings at your museum? How could evaluation use be improved?

Evaluation Practice

1. Have you conducted or commissioned a program evaluation in the previous year?

Probe (if YES)

a) Which specific programs did you evaluate? How many was that?
b) How do you decide which programs to evaluate?
c) When planning to undertake the evaluation, did you prepare a logic model/evaluation framework/results chain?
d) Were the evaluations summative or formative?
e) Who was involved in the evaluation (i.e., program participants, professional evaluator)? What were their roles?
f) Which methods and approaches did you use to evaluate the program(s)?
g) What type of information did you collect (outcomes vs. outputs)?
h) How was this information managed and analyzed? What software, if any, did you use?
i) How were the results reported (i.e., evaluation report, staff meeting)?
j) What challenges did you encounter if any?
k) How satisfied are you with the quality of the program evaluation(s) performed in the last year?

2. How frequently do you conduct evaluations of your educational programs?

3. How satisfied are you with the degree of educational program evaluation at your museum?

4. What would enable you to conduct more and higher quality evaluations of your educational programs? What kind of evaluation assistance would you benefit from?

Evaluation Capacity Building

1. Has your museum provided you with formal professional development opportunities to enhance your evaluative knowledge and skills such as courses or workshops? Explain.

2. Are there any other less formal strategies the museum has used to build the evaluation capacity of its staff (i.e., networks/partnerships)?
3. What would you say your evaluation-related professional development needs are at this time? And, what steps do you think the museum could take to meet your professional development needs?

4. What do you think are the central barriers to evaluation/visitor studies capacity at your museum? How do you think these could best be overcome?

Wrap-Up

Now, I’ve asked you a lot of questions, but I may not have asked about something that is important to you. Is there anything else that you think I should know about, but that we did not talk about?

Thank you so very much for all the information you have provided—it has helped inform my thinking around program evaluation in art museums. I enjoyed our interview!
## Appendix F

### Start List of Codes

<table>
<thead>
<tr>
<th>Context</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON: Size</td>
<td>CON-S</td>
</tr>
<tr>
<td>CON: Type</td>
<td>CON-T</td>
</tr>
<tr>
<td>CON: Age</td>
<td>CON-A</td>
</tr>
<tr>
<td>CON: Geographical Location</td>
<td>CON-GL</td>
</tr>
<tr>
<td>CON: Visitor Demographics</td>
<td>CON-VD</td>
</tr>
<tr>
<td>CON: Sources of Funding</td>
<td>CON-SoF</td>
</tr>
<tr>
<td>CON: Mandate</td>
<td>CON-M</td>
</tr>
<tr>
<td>CON: Institutional Structure</td>
<td>CON-IS</td>
</tr>
<tr>
<td>CON: Other</td>
<td>CON-O</td>
</tr>
<tr>
<td><strong>Culture</strong></td>
<td></td>
</tr>
<tr>
<td>CUL: Valuing of Education</td>
<td>CUL-Ed</td>
</tr>
<tr>
<td>CUL: Openness to Evaluation</td>
<td>CUL-Eval</td>
</tr>
<tr>
<td>CUL: Other</td>
<td>CUL-O</td>
</tr>
<tr>
<td><strong>Demand for Evaluation</strong></td>
<td></td>
</tr>
<tr>
<td>DEM: Internal Impetus</td>
<td>DEM-Internal</td>
</tr>
<tr>
<td>DEM: External Pressures</td>
<td>DEM-External</td>
</tr>
<tr>
<td>DEM: Other</td>
<td>DEM-O</td>
</tr>
<tr>
<td><strong>Resources and Supports</strong></td>
<td></td>
</tr>
<tr>
<td>R&amp;S: Human Resources</td>
<td>R&amp;S-HR</td>
</tr>
<tr>
<td>R&amp;S: Time</td>
<td>R&amp;S-T</td>
</tr>
<tr>
<td>R&amp;S: Financial Resources</td>
<td>R&amp;S-S</td>
</tr>
<tr>
<td>R&amp;S: Other Supports</td>
<td>R&amp;S-OS</td>
</tr>
<tr>
<td>R&amp;S: Technology</td>
<td>R&amp;S-T</td>
</tr>
<tr>
<td>R&amp;S: Other</td>
<td>R&amp;S-O</td>
</tr>
<tr>
<td><strong>Capacity to Do and Use Evaluation</strong></td>
<td></td>
</tr>
<tr>
<td>CAP: Knowledge and Skills</td>
<td>CAP-K&amp;S</td>
</tr>
<tr>
<td>CAP: Use of Findings</td>
<td>CAP-UoF</td>
</tr>
<tr>
<td>CAP: Process Use</td>
<td>CAP-Proc</td>
</tr>
<tr>
<td>CAP: Other</td>
<td>CAP-O</td>
</tr>
<tr>
<td><strong>Evaluation Activities</strong></td>
<td></td>
</tr>
<tr>
<td>ACT: What Gets Evaluated</td>
<td>ACT-What</td>
</tr>
<tr>
<td>ACT: Methods</td>
<td>ACT-M</td>
</tr>
<tr>
<td>ACT: Who Was Involved</td>
<td>ACT-Who</td>
</tr>
<tr>
<td>ACT: Type of Evaluation</td>
<td>ACT-ToE</td>
</tr>
<tr>
<td>ACT: Type of Information</td>
<td>ACT-ToI</td>
</tr>
<tr>
<td>ACT: Satisfaction with Frequency/Quality</td>
<td>ACT-S</td>
</tr>
<tr>
<td>ACT: Other</td>
<td>ACT-O</td>
</tr>
<tr>
<td><strong>Evaluation Capacity Building</strong></td>
<td></td>
</tr>
<tr>
<td>ECB: Direct Activities</td>
<td>ECB-D</td>
</tr>
<tr>
<td>Evaluation Capacity Building</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>ECB: Indirect Activities</td>
<td>ECB-I</td>
</tr>
<tr>
<td>ECB: Other</td>
<td>ECB-O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenges</td>
<td>CH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Needs</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needs</td>
<td>N</td>
</tr>
</tbody>
</table>
Appendix G
Sample Matrix Coding Charts

For Phase One Interviews

<table>
<thead>
<tr>
<th>Transcript</th>
<th>DEM-II</th>
<th>DEM-EP</th>
<th>DEM-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Demand for Evaluation: Internal Impetus = DEM-II
Demand for Evaluation: External Pressures = DEM-EP
Demand for Evaluation: Challenges = DEM-C

For Phase Two Case Studies

<table>
<thead>
<tr>
<th>ECB</th>
<th>Case</th>
<th>ECB-A</th>
<th>ECB-C</th>
<th>ECB-N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ECB Activities = ECB-A
ECB Challenges = ECB-C
ECB Needs = ECB-N
Appendix H

Phase Two Combined Letter of Information and Consent Form

For Executive Directors:

Study Title: Painting a Portrait of Program Evaluation in Canadian Art Museums
Name of Student Researcher: Agnieszka Chalas, Queen’s University

Name of Supervisor: Chris DeLuca, Queen’s University

I am Agnieszka Chalas, a doctoral candidate in the Faculty of Education at Queen’s University working under the supervision of Dr. Chris DeLuca. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in Canada most active in evaluation and, as a result, was selected for in-depth case study. This phase in my research is designed to take a closer look at program evaluation and visitor research at these museums and draw out specific examples of institutional successes (i.e., exemplary program evaluation, mainstreaming, and capacity building practices, etc.). There are no known risks to your participation. The study will benefit art museum practitioners by providing them with examples of best practices and a practical tool for both reflecting on and building their capacity for evaluation. The study is funded by the Social Sciences and Humanities Research Council of Canada.

I will be inviting up to ten participants at your museum to participate in this case study. If you agree to be one of them, I will interview you for approximately 30-45 minutes at your place of work at a time that is convenient for you. With your permission, the interview will be audio-recorded and later transcribed. During the interview, I also hope to collect any relevant documents that may add to my understanding of evaluation/visitor research at your museum. I may also ask you to (a) identify additional staff members who might be good sources of information about program evaluation/visitor studies at your museum and (b) any relevant evaluation/research activities that I may observe.

Once I have analyzed the interview, document, and observational data, I will draft a case report for your museum which I will share with you and those of your colleagues who also participated in the study. At such a time, you will be encouraged to review and provide feedback on the report. I will then finalize your case report based on the feedback received and will use it to as a basis for cross-case analysis (i.e., between the other two Canadian art museums also identified as exemplars).

Your participation in the study is voluntary—you don’t have to answer any questions you don’t want to, can stop participating in the interview at any time, and are free to withdraw from the study up until three months following the interview by contacting me at a.chalas@queensu.ca. If you choose to withdraw, you may request removal of all or part of your data from the study. In accordance with the Faculty of Education’s policy, your data will be retained for a minimum of five years on a password protected computer and no one other than me will have access to any of the data.
The findings from this study may be published in academic journals and presented at professional conferences. Because this study pinpoints your art museum as an ‘exemplar’ and confidentiality is not a desirable choice in case study research, I would like to **accurately identify the names of the museums selected for case study as well as participants’ names and job titles** in any such publications/presentations. This kind of disclosure would ensure that important cultural and contextual information about your museum is not eliminated from this phase of my study. Should you be uncomfortable with full disclosure, the attached consent form presents you with other options around disclosure including the option not to be identified.

If you have any questions about this study, please contact me, Agnieszka Chalas, at a.chalas@queensu.ca / 613-876-5427, or my supervisor Dr. Chris DeLuca at cdeluca@queensu.ca / 613-533-6000 ext. 77675. Any ethical concerns about the study may be directed to the General Research Ethics Board (GREB) at 1-844-535-2988 (Toll free in North America) or chair.GREB@queensu.ca.

Sincerely,

[Signature]

Agnieszka Chalas, Ph.D. Candidate, Faculty of Education, Queen’s University

Keep one copy of the Letter of Information and Consent Form for your records and return one copy to the researcher, Agnieszka Chalas, a.chalas@queensu.ca.

The above Letter of Information provides you with the necessary details to help you make an informed choice about participation. All your questions should be answered to your satisfaction before you decide whether or not to participate in this research study.

Please check the box that most accurately describes your level of consent:

- I consent to the disclosure of the name of the museum at which I am the Executive Director (or equivalent)

OR

- I would prefer if the name of the museum at which I am the Executive Director (or equivalent) were replaced with a pseudonym

- I consent to the disclosure of my name and job title

OR

- I consent to the disclosure of my job title only
I would prefer if my identity were protected to the extent possible. I understand that, given the small number of art museums that fit the study criteria, my identity may be deduced by others.

By signing below, I am verifying that: I have read the Letter of Information, agree to participate in the study, and all of my questions have been answered.

Name: __________________________________________
Signature: _______________________________________
Date: ________________________________

I wish to receive a copy of the research findings!

For prospective participants whose Executive Directors have elected to disclose the names of the museums at which they work:

Study Title: Painting a Portrait of Program Evaluation in Canadian Art Museums
Name of Student Researcher: Agnieszka Chalas, Queen’s University
Name of Supervisor: Chris DeLuca, Queen’s University

I am Agnieszka Chalas, a doctoral candidate in the Faculty of Education at Queen’s University working under the supervision of Dr. Chris DeLuca. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in Canada most active in evaluation and, as a result, was selected for in-depth case study. This phase in my research is designed to take a closer look at program evaluation and visitor research at these museums and draw out specific examples of institutional successes (i.e., exemplary program evaluation, mainstreaming, and capacity building practices, etc.). There are no known risks to your participation. The study will benefit art museum practitioners by providing them with examples of best practices and a practical tool for both reflecting on and building their capacity for evaluation. The study is funded by the Social Sciences and Humanities Research Council of Canada.

I will be inviting up to ten participants at your museum to participate in this case study. If you agree to be one of them, I will interview you for approximately 30-45 minutes at your place of work at a time that is convenient for you. With your permission, the interview will be audio-recorded and later transcribed. During the interview, I also hope to collect any relevant documents that may add to my understanding of evaluation/visitor research at your museum. I may also ask you to (a) identify additional staff members who might be good sources of information about program evaluation/visitor studies at your museum and (b) any relevant evaluation/research activities that I may observe.

Once I have analyzed the interview, document, and observational data, I will draft a case report for your museum which I will share with you and those of your colleagues who also participated in the study. At such a time, you will be encouraged to review and provide feedback on the report. I will then finalize your case report based on the feedback received and will use it to as a basis for cross-case analysis (i.e., between the other two Canadian art museums also identified as exemplars).
Your participation in the study is voluntary—you don’t have to answer any questions you don’t want to, can stop participating in the interview at any time, and are free to withdraw from the study up until three months following the interview by contacting me at a.chalas@queensu.ca. If you choose to withdraw, you may request removal of all or part of your data from the study. In accordance with the Faculty of Education’s policy, your data will be retained for a minimum of five years on a password protected computer and no one other than me will have access to any of the data.

The findings from this study may be published in academic journals and presented at professional conferences. Because this study pinpoints your art museum as an ‘exemplar’, the Executive Director of the museum at which you work has chosen to accurately identify your museum in any such publications/presentations. The attached consent form, presents you with different options around disclosure. Please bear in mind the fact that your museum’s name will be made known to others when filling out the form.

If you have any questions about this study, please contact me, Agnieszka Chalas, at a.chalas@queensu.ca / 613-876-5427, or my supervisor Dr. Chris DeLuca at cdeluca@queensu.ca / 613-533-6000 ext. 77675. Any ethical concerns about the study may be directed to the General Research Ethics Board (GREB) at 1-844-535-2988 (Toll free in North America) or chair.GREB@queensu.ca.

Sincerely,

Agnieszka Chalas, Ph.D. Candidate, Faculty of Education, Queen’s University

Keep one copy of the Letter of Information and Consent Form for your records and return one copy to the researcher, Agnieszka Chalas, a.chalas@queensu.ca.

The above Letter of Information provides you with the necessary details to help you make an informed choice about participation. All your questions should be answered to your satisfaction before you decide whether or not to participate in this research study.

Please check the box that most accurately describes your level of consent:

I consent to the disclosure of my name and job title

OR

I consent to the disclosure of my job title only

I would prefer if my identity were protected to the extent possible. I understand that, given the small number of art museums that fit the study criteria and the fact that my Executive Director
I have chosen to disclose the name of the museum at which I work, my identity may be deduced by others.

By signing below, I am verifying that: I have read the Letter of Information, agree to participate in the study, and all of my questions have been answered.

Name: ______________________________
Signature: ___________________________
Date: ______________________________

I wish to receive a copy of the research findings!

For prospective participants whose Executive Directors have elected to replace the name of the museum at which they work with a pseudonym:

Study Title: Painting a Portrait of Program Evaluation in Canadian Art Museums
Name of Student Researcher: Agnieszka Chalas, Queen’s University
Name of Supervisor: Chris DeLuca, Queen’s University

I am Agnieszka Chalas, a doctoral candidate in the Faculty of Education at Queen’s University working under the supervision of Dr. Chris DeLuca. Based on the results of a pan-Canadian interview study, your museum was identified as one of two art museums in Canada most active in evaluation and, as a result, was selected for in-depth case study. This phase in my research is designed to take a closer look at program evaluation and visitor research at these museums and draw out specific examples of institutional successes (i.e., exemplary program evaluation, mainstreaming, and capacity building practices, etc.). There are no known risks to your participation. The study will benefit art museum practitioners by providing them with examples of best practices and a practical tool for both reflecting on and building their capacity for evaluation. The study is funded by the Social Sciences and Humanities Research Council of Canada.

I will be inviting up to ten participants at your museum to participate in this case study. If you agree to be one of them, I will interview you for approximately 30-45 minutes at your place of work at a time that is convenient for you. With your permission, the interview will be audio-recorded and later transcribed. During the interview, I also hope to collect any relevant documents that may add to my understanding of evaluation/visitor research at your museum. I may also ask you to (a) identify additional staff members who might be good sources of information about program evaluation/visitor studies at your museum and (b) any relevant evaluation/research activities that I may observe.

Once I have analyzed the interview, document, and observational data, I will draft a case report for your museum which I will share with you and those of your colleagues who also participated in the study. At such a time, you will be encouraged to review and provide feedback on the report. I will then finalize your case report based on the feedback received and will use it to as a basis for cross-case analysis (i.e., between the other two Canadian art museums also identified as exemplars).
Your participation in the study is voluntary—you don’t have to answer any questions you don’t want to, can stop participating in the interview at any time, and are free to withdraw from the study up until three months following the interview by contacting me at a.chalas@queensu.ca. If you choose to withdraw, you may request removal of all or part of your data from the study. In accordance with the Faculty of Education’s policy, your data will be retained for a minimum of five years on a password protected computer and no one other than me will have access to any of the data.

The findings from this study may be published in academic journals and presented at professional conferences. However, the name of the museum at which you work will be replaced with a pseudonym in any such publications/presentations. The attached consent form presents you with different options around disclosure. Please bear this in mind when filling out the form.

If you have any questions about this study, please contact me, Agnieszka Chalas, at a.chalas@queensu.ca / 613-876-5427, or my supervisor Dr. Chris DeLuca at cdeluca@queensu.ca / 613-533-6000 ext. 77675. Any ethical concerns about the study may be directed to the If you have any ethical concerns General Research Ethics Board (GREB) at 1-844-535-2988 (Toll free in North America) or chair.GREB@queensu.ca.

Sincerely,

Agnieszka Chalas, Ph.D. Candidate, Faculty of Education, Queen’s University

Keep one copy of the Letter of Information and Consent Form for your records and return one copy to the researcher, Agnieszka Chalas, a.chalas@queensu.ca.

The above Letter of Information provides you with the necessary details to help you make an informed choice about participation. All your questions should be answered to your satisfaction before you decide whether or not to participate in this research study.

Please check the box that most accurately describes your level of consent:

I consent to the disclosure of my job title

OR

I would prefer if my identity were protected to the extent possible. I understand that, given the small number of art museums that fit the study criteria, my identity may be deduced by others

By signing below, I am verifying that: I have read the Letter of Information, agree to participate in the study, and all of my questions have been answered.
Name: ______________________________
Signature: __________________________
Date: ______________________________

I wish to receive a copy of the research findings!
Appendix I
Phase Two Interview Question Matrix

Sample Questions for Interviews with Various Case Study Participants

Executive Director = ED
Education Staff = EdS
Exhibitions Staff = ExS
Visitor Studies Staff (includes Marketing) = VSS

<table>
<thead>
<tr>
<th>Context</th>
<th>Informant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is your museum’s educational mission?</td>
<td>Eds</td>
</tr>
<tr>
<td>What is the full extent of the educational and public programs offered by your department (i.e., how many, target audiences) and how are they funded?</td>
<td>Eds</td>
</tr>
<tr>
<td>Can you describe for me the composition of your educational department (i.e., number and role of staff)?</td>
<td>Eds</td>
</tr>
<tr>
<td>What is your department’s annual operating budget?</td>
<td>Eds</td>
</tr>
<tr>
<td>What would you say are the museum’s institutional priorities?</td>
<td>ED, Eds</td>
</tr>
<tr>
<td>Can you describe for me the museum’s organizational structure?</td>
<td>ED, Eds</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Culture</th>
<th>Informant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is your museum subject to any external federal or provincial accountability policies such as the Treasury Board of Canada Secretariat Federal Accountability Act for example?</td>
<td>ED</td>
</tr>
<tr>
<td>Does your museum have a broad-based evaluation agenda or an evaluation plan in place?</td>
<td>ED, Eds</td>
</tr>
<tr>
<td>How integral do you think evaluation is to your operations and internal accountability framework?</td>
<td>ED, Eds</td>
</tr>
<tr>
<td>Is evaluation linked to broader management systems like strategic planning?</td>
<td>ED, Eds</td>
</tr>
<tr>
<td>How do you or the Board support visitor studies or program evaluation at your museum?</td>
<td>ED</td>
</tr>
<tr>
<td>What kinds of opportunities do you provide staff to enhance their evaluative competencies?</td>
<td>ED</td>
</tr>
<tr>
<td>To what degree are you interested in institutionalizing or broadening the role of evaluation at the museum?</td>
<td>ED</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand for Evaluation</th>
<th>Informant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there an internal mandate for evaluation information at the museum?</td>
<td>ED</td>
</tr>
<tr>
<td>What was the main reason your last evaluation was conducted? Was it internally or externally driven?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Which percentage of funders require evaluation? How many different funders are you accountable to?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Do your funders require a formal evaluation plan, specify the types of evaluation information they want you to provide or methods they want you to use or who should be responsible for conducting the evaluation?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What type and amount of information do funders require and to what extent is this information that you would not otherwise have collected?</td>
<td>EdS, ExS, VSS</td>
</tr>
</tbody>
</table>
Table Continued

<table>
<thead>
<tr>
<th>Demand for Evaluation</th>
<th>Informant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you perceive the evaluation expectations of funders?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Do you find meeting funders’ evaluation demands challenging?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How could funders better support you in meeting their accountability demands?</td>
<td>EdS, ExS, VSS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resources and Supports</th>
<th>Informant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What percentage of the museum’s revenue is allocated to research and evaluation?</td>
<td>ED</td>
</tr>
<tr>
<td>What percentage of your time is dedicated to evaluation?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>To what extent do your funders provide financial resources or allow project funds to be used for evaluation purposes?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>What percentage of a grant total is dedicated to supporting evaluation activities?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>To what degree do funders provide other supports (i.e., access to evaluation training or funding for training, advice on evaluation, technical assistance)?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>What kind of evaluation support do you receive (financial or otherwise from your museum’s leadership? Is there an internal evaluation budget in place?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>Are you provided with opportunities to enhance your evaluative skills?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>What kind of software do you use to compile, analyze, and manage evaluation information?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>What are some of the barriers you face to evaluating your education programs?</td>
<td>Eds, ExS, VSS</td>
</tr>
<tr>
<td>What would you say your evaluation resource needs are at this time? What kind of supports do you require from funders and leadership to effectively evaluate your programs?</td>
<td>Eds, ExS, VSS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capacity to Do and Use Evaluation</th>
<th>Informant(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe your background for me. Did you receive any training in evaluation as part of your professional preparation?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Tell me about your previous practical visitor research/evaluation experience (i.e., number of evaluations, years of experience, working alongside an evaluator)</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What do you think is the greatest strength you bring to the evaluation of museum education programs?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How comfortable do you feel with conducting program evaluations (i.e., framing research and evaluation questions; planning evaluation studies; addressing ethical issues in research; designing instruments; collecting, analyzing, and interpreting data; reporting evaluation findings)?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How comfortable do you feel with conducting visitor studies (i.e., exhibit exit surveys, institutional exit surveys, tracking studies, sweep rates, visitor satisfaction surveys, market research etc.)?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How do you use the evaluation/visitor studies information you collect? For what purposes?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How are findings shared across the museum and/or disseminated beyond the museum? Who reviews the studies internally?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>To what degree do research and evaluation findings feed into strategic planning?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>To what degree are study recommendations acted upon?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>To what degree do you use research and evaluation findings for decision making purposes?</td>
<td>EdS, ExS, VSS</td>
</tr>
</tbody>
</table>
Table Continued

<table>
<thead>
<tr>
<th><strong>Capacity to Do and Use Evaluation</strong></th>
<th><strong>Informant(s)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>What did you learn from your last evaluation/visitor study?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>To what degree do you think staff attitudes have changed or their evaluation skills have increased as a result of participating in an evaluation/visitor study?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What would you say are the biggest barriers to conducting research and evaluation studies and using the findings of such studies at your museum?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What would you say your research and evaluation related professional development needs are at this time?</td>
<td>EdS, ExS, VSS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Evaluation Activities</strong></th>
<th><strong>Informant(s)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you conducted or commissioned an evaluation/visitor study in the previous year? How many was that?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What specifically did you evaluate/study?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How do you decide what to evaluate/study?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>When planning to undertake the study, did you prepare a logic model/evaluation framework/results chain?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Was the study summative or formative?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Who was involved in the study and what were their roles?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Which evaluation/research methods and approaches did you use?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What type of information did you collect?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How was this information managed and analyzed? What software, if any, did you use?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How were the results reported?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What challenged did you encounter if any?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How satisfied are you with the quality of the study? How satisfied are you with the quality and quantity of the research and evaluation being conducted at your museum?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>How frequently do you conduct research and evaluation studies?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What would enable you to conduct more and higher quality evaluations of your educational programs? What specific kinds of evaluation assistance would you benefit from?</td>
<td>EdS, ExS, VSS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>ECB</strong></th>
<th><strong>Informant(s)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe for me any strategies or activities the museum has used to build the evaluation capacity of its staff</td>
<td>ED</td>
</tr>
<tr>
<td>Has your museum provided you with formal professional development opportunities to enhance your evaluative knowledge and skills such as courses or workshops? Explain</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>Can you think of any other less formal strategies the museum has used to build the evaluation capacity of its staff (e.g., networks/partnerships)?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What would you say your evaluation-related professional development needs are at this time? And, what steps do you think the museum could take to meet your professional development needs?</td>
<td>EdS, ExS, VSS</td>
</tr>
<tr>
<td>What do you think are the main barriers to research and evaluation at your museum? How do you think these could best be overcome?</td>
<td>EdS, ExS, VSS</td>
</tr>
</tbody>
</table>
## Appendix J

### Final List of Codes Applied During Phase One

<table>
<thead>
<tr>
<th>Context</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON: Size</td>
<td>CON-S</td>
<td>5</td>
<td>I would say federal and larger budget museums they have the luxury of doing more evaluation.</td>
</tr>
<tr>
<td>CON: Type</td>
<td>CON-T</td>
<td>7</td>
<td>Art museums in general are behind the curve when it comes to doing visitor research and evaluation and really using it and trying to inform their exhibition and program development process.</td>
</tr>
<tr>
<td>CON: Age/Maturity</td>
<td>CON-A</td>
<td>1</td>
<td>Since we’ve only been operating for two years we haven’t had our exhibitions spaces open to the public.</td>
</tr>
<tr>
<td>CON: Geographical Location/Accountability</td>
<td>CON-GL/AC</td>
<td>3</td>
<td>In art museums in the States evaluation is more embedded in their practice. If you look at the Gander Museum of Art or the Detroit Institute of Art I think visitor studies and evaluation are built in to their operations.</td>
</tr>
<tr>
<td>CON: Institutional Structure</td>
<td>CON-Struc</td>
<td>1</td>
<td>The old school traditional hierarchies still definitely exist here in terms of the internal curatorial kind of rule.</td>
</tr>
<tr>
<td>CON: Institutional Stability</td>
<td>CON-Stab</td>
<td>17</td>
<td>I think because of our challenges we’ve been in kind of a renovation/construction mode for a while and we had to move programs off-site.</td>
</tr>
<tr>
<td>Culture</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>CUL: Valuing of Education</td>
<td>CUL-Ed</td>
<td>24</td>
<td>I mean [education] is highly valued but we’re not necessarily part of the decision making of what goes on in the museum. So, we’re valued as long as we continue to bring in students but not necessarily as a voice with the designers, or curators, or things like that yet.</td>
</tr>
<tr>
<td>CUL: Openness to Evaluation</td>
<td>CUL-Eval</td>
<td>60</td>
<td>I think it is valued and I think evaluation is a helpful tool to communicate the value of educational programming. . . I can say that at [gallery] there is definitely a desire and an understanding that we have to be visitor-centred in our approach.</td>
</tr>
</tbody>
</table>
### Table Continued

<table>
<thead>
<tr>
<th>Culture</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL: Learning Orientation</td>
<td>CUL-LO</td>
<td>3</td>
<td>We’re a risk-adverse institution</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM: Internal Impetus</td>
<td>DEM-I</td>
<td>11</td>
<td>It’s never been asked of us but I know if all of a sudden there was no evaluation happening there would be questions asked but I think because evaluation has always been there</td>
</tr>
<tr>
<td>DEM: External Pressure</td>
<td>DEM-E</td>
<td>19</td>
<td>Yes. definitely. There’s one funder in particular that asks you to, almost like a school curriculum rubric where they kind of want you to predict outcomes, say how you were planning to meet those outcomes, and what you believe those outcomes were</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resources/Supports</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;S: Human Resources</td>
<td>R&amp;S-HR</td>
<td>30</td>
<td>[Name] is the Manager, Interpretation and Visitor Research and there are two employed people doing all the data collection</td>
</tr>
<tr>
<td>R&amp;S: Time</td>
<td>R&amp;S-T</td>
<td>12</td>
<td>Well we’ve been pretty good at [evaluation] lately but I would say less than 10% of my time</td>
</tr>
<tr>
<td>R&amp;S: Financial Resources</td>
<td>R&amp;S-S</td>
<td>12</td>
<td>They do [provide us with financial resources for evaluation]. I think it can be up to 10% specifically for evaluation</td>
</tr>
<tr>
<td>R&amp;S: Other Supports</td>
<td>R&amp;S-O</td>
<td>8</td>
<td>Every year they run an information seminar but it’s also largely directed at evaluation so they want to help you figure out how to write your grant but also how to report on it</td>
</tr>
<tr>
<td>R&amp;S: Technology</td>
<td>R&amp;S-Tech</td>
<td>9</td>
<td>Our new point of sale system has some built in evaluation help . . . through its report compiler</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP: Knowledge and Skills</td>
<td>CAP-K&amp;S</td>
<td>35</td>
<td>I did a lot of statistics and methodology when I did a Museum Studies degree at [university] and my main focus was visitor research</td>
</tr>
<tr>
<td>CAP: Instrumental Use</td>
<td>CAP-I</td>
<td>33</td>
<td>One of the bigger changes that we have made recently is to move our March Break program from being a half-day to a full-day event only</td>
</tr>
<tr>
<td>Capacity</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CAP: Conceptual Use</td>
<td>CAP-C</td>
<td>11</td>
<td>We have published one journal article outlining those results and another is pending</td>
</tr>
<tr>
<td>CAP: Persuasive Use</td>
<td>CAP-Per</td>
<td>3</td>
<td>There is a bit of an advocacy angle that goes into [evaluation]—to demonstrate the value of programs to visitors and to demonstrate the value of evaluation to senior management</td>
</tr>
<tr>
<td>CAP: Non-Use</td>
<td>CAP-N</td>
<td>3</td>
<td>I think that there can be a tendency [in art museums] like in a lot of places to gather a lot of information but not really look at it or analyze it or act on it</td>
</tr>
<tr>
<td>CAP: Process Use</td>
<td>CAP-Proc</td>
<td>5</td>
<td>I guess from working with [hospital] has come an awareness of the kinds of tools and the kinds of questions that you look for so that’s something that I can take to looking at other projects as well. And it becomes capacity building for both sides of those involved in a partnership in ways that I had not originally anticipated</td>
</tr>
<tr>
<td><strong>Activities</strong></td>
<td><strong>Abbreviation</strong></td>
<td><strong>Frequency</strong></td>
<td><strong>In-Text Example</strong></td>
</tr>
<tr>
<td>ACT: Type of Evaluation</td>
<td>ACT-TOE</td>
<td>27</td>
<td>For the (program) we didn’t do front-end evaluation, we did summative evaluation</td>
</tr>
<tr>
<td>ACT: What Gets Evaluated</td>
<td>ACT-What</td>
<td>56</td>
<td>We’ve managed to pretty much evaluate every single program including the docent training program which I do</td>
</tr>
<tr>
<td>ACT: Methods</td>
<td>ACT-M</td>
<td>63</td>
<td>We timed visitors through the (space) to be able to get an average length of visit, we also calculated the sweep rate index of the (space), our security guards tracked attendance and both front-line staff and myself interviewed 53 visitors using a conversational format</td>
</tr>
<tr>
<td>ACT: Type of Information Collected</td>
<td>ACT-TOI</td>
<td>50</td>
<td>The only things that we really look at are things like numbers of participants and revenue generated</td>
</tr>
<tr>
<td>ACT: Who Was Involved</td>
<td>ACT-Who</td>
<td>48</td>
<td>We do a lot of consulting with groups. We do a lot of consulting with artists and art historians around messaging and that kind of thing</td>
</tr>
<tr>
<td>Activities</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td>ACT: Satisfaction</td>
<td>ACT-S</td>
<td>16</td>
<td>I don’t think we do [evaluation] nearly enough as an education department</td>
</tr>
<tr>
<td>ECB: Direct Activities</td>
<td>ECB-D</td>
<td>8</td>
<td>We just learned about basic evaluation like what’s the difference between front-end evaluation and the other kinds and all that kind of stuff and so really it was sort of a crash course on evaluation and visitor studies</td>
</tr>
<tr>
<td>ECB: Indirect Activities</td>
<td>ECB-I</td>
<td>19</td>
<td>Well, we do have the (local museum educators’ group) and we have had sessions on evaluation there</td>
</tr>
<tr>
<td>Challenges</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>CH: Lack of Resources and Supports</td>
<td>CH-R&amp;S</td>
<td>33</td>
<td>There isn’t the time and there aren’t the resources</td>
</tr>
<tr>
<td>CH: Lack of Internal Demand/Commitment</td>
<td>CH-ID/C</td>
<td>7</td>
<td>Unless someone gives me capacity and prioritizes evaluation it’s not happening</td>
</tr>
<tr>
<td>CH: Practice-Based Challenges</td>
<td>CH-Prac</td>
<td>13</td>
<td>We’ve tried personal interceptions but it’s loud and there’s always loud music and its very much a party environment so it’s really hard to get people to participate in one-on-one interviews</td>
</tr>
<tr>
<td>CH: Marginalization of Education</td>
<td>CH-Ed</td>
<td>6</td>
<td>There are big programs that I think should be evaluated because they’re big elements of the gallery’s operating but we don’t think of them that way because they operate at arms-length from the exhibitions</td>
</tr>
<tr>
<td>CH-Other</td>
<td>CH-O</td>
<td>6</td>
<td>It’s really hard to train people without museum evaluation experience</td>
</tr>
<tr>
<td>Needs</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>N: Resources &amp; Supports</td>
<td>N-R&amp;S</td>
<td>28</td>
<td>I think we could do more if we had a dedicated position for evaluation and like at the Museum of Science and Technology who oversees evaluation not just for education but kind of across the board</td>
</tr>
<tr>
<td>N: Internal Demand for and Organizational Commitment to Institutionalizing Evaluation</td>
<td>N-ID/OC</td>
<td>9</td>
<td>The drive for evaluation should be really be coming from all levels in the institution from the Board to the Director and management teams to the workers and it should be part of the strategic plan, part of what we do in all departments</td>
</tr>
<tr>
<td>Needs</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>N: External Demand for Evaluation</td>
<td>N-ED</td>
<td>2</td>
<td>A requirement from funders would in my opinion be the best way to ensure that evaluation happens</td>
</tr>
<tr>
<td>N: Training in Evaluation</td>
<td>N-T</td>
<td>15</td>
<td>Sometimes I’d be looking at in these reports and these pie charts that Survey Monkey would spit out and trying to analyze the data and that was a real struggle for me really. There’s a lot of opportunity for me to get training there.</td>
</tr>
<tr>
<td>Reactivity/Opportunity</td>
<td>Abbreviation</td>
<td>Frequency</td>
<td>In-Text Example</td>
</tr>
<tr>
<td>Reactivity</td>
<td>REAC</td>
<td>11</td>
<td>This is making me think that what I’m going to try to do is find some money for next year to hire an educator to evaluate our programs.</td>
</tr>
<tr>
<td>Opportunities</td>
<td>OPP</td>
<td>4</td>
<td>There’s a park in San Diego with a number of small museums who kind of group their resources together to do audience evaluation—Balboa Park I think it is. They probably don’t have the resources to conduct evaluations independently but if a number of them get together they can.</td>
</tr>
</tbody>
</table>
# Appendix K

## Final List of Codes Applied During Phase Two

### To Interview Transcripts Associated with Case Portrait 1: The Montreal Museum of Fine Arts

<table>
<thead>
<tr>
<th>Context</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON: Geographical Location/Accountability Context</td>
<td>CON-GL/AC</td>
<td>4</td>
<td>They are my only benchmark in Canada. Then if I want to benchmark outside of Canada I look to the Boston Museum of Art but they have budgets that are so up there that it’s not the same situation. It’s very, very different</td>
</tr>
<tr>
<td>CON: Sources of Funding</td>
<td>CON-SoF</td>
<td>5</td>
<td>For example our wellness programs are supported by the Rossy Family</td>
</tr>
<tr>
<td>CON: Institutional Structure</td>
<td>CON-Struc</td>
<td>3</td>
<td>So, it’s like we work together but there are 3 marketing heads at the museum with three different mandates and the mandates really are intertwined but still it’s different aspects of marketing at the museum</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Culture</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL: Valuing of Education</td>
<td>CUL-Ed</td>
<td>11</td>
<td>I think it is one of the essential missions of our museum . . . for our director the importance of education seems to be a priority of hers</td>
</tr>
<tr>
<td>CUL: Openness to Evaluation</td>
<td>CUL-Eval</td>
<td>11</td>
<td>Well, it’s essential. In order to get the feedback directly from the people who are using the programs . . . the only way you are going to improve a program is by listening to the people</td>
</tr>
<tr>
<td>CUL: Learning Orientation</td>
<td>CUL-LO</td>
<td>3</td>
<td>And it’s a culture that is tends to be progressive and dynamic and evolving where staff are updating themselves, re-educating themselves</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM: Internal Impetus</td>
<td>DEM-I</td>
<td>5</td>
<td>I would say that in any museum that I know of most of the time big data is more or less always driven by the membership department because that’s our core business</td>
</tr>
<tr>
<td>DEM: External Pressure</td>
<td>DEM-E</td>
<td>6</td>
<td>You know it’s not there that the demand is the biggest it’s mostly for sponsorship that they ask for very specific numbers</td>
</tr>
</tbody>
</table>
### Resources/Supports

<table>
<thead>
<tr>
<th>Resources/Supports</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;S: Human Resources</td>
<td>R&amp;S-HR</td>
<td>18</td>
<td>I’m working as well with a media firm that does all the media planning for us. They have the tools to analyze the postal codes</td>
</tr>
<tr>
<td>R&amp;S: Time</td>
<td>R&amp;S-T</td>
<td>1</td>
<td>I don’t think I’d be able to answer that question I really don’t know… I really have no idea how much time I spend</td>
</tr>
<tr>
<td>R&amp;S: Technology</td>
<td>R&amp;S-T</td>
<td>4</td>
<td>The system that we use across the museum for everything . . . every teacher that has ever reserved a school visit their information is in there—the times, the dates, the number of students, the number people that have come with them, everything is there</td>
</tr>
</tbody>
</table>

### Capacity

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP: Knowledge and Skills</td>
<td>CAP-K&amp;S</td>
<td>7</td>
<td>I went on to do a Masters in Educational Psychology where my emphasis was on program evaluation</td>
</tr>
<tr>
<td>CAP: Instrumental Use</td>
<td>CAP-I</td>
<td>25</td>
<td>In future if I hire Flanny again and I likely will I will say in the program description that she does it only in French and not bilingually. That’s what that woman mentioned in the evaluation: “in the program description you should say that’s it’s only in French”</td>
</tr>
<tr>
<td>CAP: Conceptual Use</td>
<td>CAP-C</td>
<td>18</td>
<td>and I just finished a draft of a chapter that tells the story of the museum’s art therapy program so we are building towards producing our own literature</td>
</tr>
<tr>
<td>CAP: Persuasive Use</td>
<td>CAP-Per</td>
<td>5</td>
<td>Knowing that the funding from [foundation] was coming to an end we wanted to gather the information first of all to show them that it was a program that was very well received by the public and at the same time we needed to gather the information in order to show future funders that “yes this program really must continue because so many people now have been using it and have come to depend on it</td>
</tr>
</tbody>
</table>

### Activities

<table>
<thead>
<tr>
<th>Activities</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT: Type of Evaluation</td>
<td>ACT-TOE</td>
<td>7</td>
<td>All of us do it very informally but we still need to have feedback in order to improve the program</td>
</tr>
</tbody>
</table>
Table Continued

<table>
<thead>
<tr>
<th>Activities</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT: What Gets Evaluated</td>
<td>ACT-What</td>
<td>13</td>
<td>In general it’s exhibitions and the annual ball, or very specific evenings for millennials etcetera</td>
</tr>
<tr>
<td>ACT: Methods</td>
<td>ACT-M</td>
<td>34</td>
<td>I have a questionnaire for group leaders and I have another one for the participants</td>
</tr>
<tr>
<td>ACT: Type of Information Collected</td>
<td>ACT-TOI</td>
<td>21</td>
<td>We ask where they come from, how many times they come to the museum, what’s their salary, what’s their educational level</td>
</tr>
<tr>
<td>ACT: Who Was Involved</td>
<td>ACT-Who</td>
<td>25</td>
<td>Lately we’ve been partnering with health agencies and hospitals like the Douglas hospital</td>
</tr>
<tr>
<td>ACT: Satisfaction</td>
<td>ACT-S</td>
<td>5</td>
<td>Right now what we are doing is fine. Most museums are not even at our level because they lack either the competence or the data base</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ECB</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECB: Indirect Activities</td>
<td>ECB-I</td>
<td>1</td>
<td>I did attend a Visitor Studies conference quite a few years ago and was very impressed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH: Lack of Resources and Supports</td>
<td>CH-R&amp;S</td>
<td>14</td>
<td>A few years ago we had evaluated our community programs. We considered doing it with an outside consultant but it would have been too costly actually we just couldn’t afford it</td>
</tr>
<tr>
<td>CH: Practice-Based Challenges</td>
<td>CH-Prac</td>
<td>1</td>
<td>I very rarely get comments from the participants because they’re very vulnerable populations. And because the brunt of my programs serve vulnerable people it’s a major issue for me</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Needs</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>N: Resources &amp; Supports</td>
<td>N-R&amp;S</td>
<td>4</td>
<td>My wish the database would be either changed or upgraded and there would be a real data analyst at the museum working under me</td>
</tr>
<tr>
<td>N: Training in Evaluation</td>
<td>N-T</td>
<td>4</td>
<td>I feel under-qualified to provide more empowered-oriented opportunities for feedback but I want to learn how to apply those approaches this context</td>
</tr>
</tbody>
</table>
Table Continued

<table>
<thead>
<tr>
<th>Reactivity/Oppportunity</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunities</td>
<td>OPP</td>
<td>1</td>
<td>[Arts-based research]—that’s the stuff that I love! I think there is a bridge that can be made between the user experience and the fine art museum that has a creative voice to it</td>
</tr>
</tbody>
</table>

To Interview Transcripts Associated with Case Portrait 2: The Art Gallery of Ontario

<table>
<thead>
<tr>
<th>Context</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON: Type</td>
<td>CON-T</td>
<td>1</td>
<td>Art museums in general are behind the curve when it comes to doing visitor research and really using it and trying to inform their exhibition development process</td>
</tr>
<tr>
<td>CON: Geographical Location/Accountability</td>
<td>CON-GL/AC</td>
<td>1</td>
<td>The big push in the States is IMLS [Institute of Museum and Library Services] and the NSF [National Science Foundation] require 10% of any grant proposal to be dedicated to evaluation. So, it’s really healthily funded in the States. It’s not funded here</td>
</tr>
<tr>
<td>CON: Institutional Structure</td>
<td>CON-Struc</td>
<td>3</td>
<td>Our education department is now called Public Programming and Learning. We’re four departments within that division. One department is interpretive planning and visitor research. Keri is in the education department or what used to be the education department. Then there’s adult programs. Then there’s young audiences. Then there’s special projects.</td>
</tr>
<tr>
<td>CON: Institutional Stability</td>
<td>CON-Stab</td>
<td>6</td>
<td>So, our department has gone through several restructures in the last 3 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Culture</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL: Valuing of Education</td>
<td>CUL-Ed</td>
<td>2</td>
<td>If you ask most people what is our biggest product they will say the exhibitions</td>
</tr>
<tr>
<td>CUL: Openness to Evaluation</td>
<td>CUL-Eval</td>
<td>16</td>
<td>I absolutely think [program evaluation] is really important and instructive. It allows us to learn how well our programs are working</td>
</tr>
</tbody>
</table>
### Table Continued

<table>
<thead>
<tr>
<th>Culture</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUL: Learning Orientation</td>
<td>CUL-LO</td>
<td>2</td>
<td>It was a way to try and shift the culture of the organization so we’re not afraid of data. So, it was like “let’s work together to try and increase data capture for example”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demand</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM: Internal Impetus</td>
<td>DEM-I</td>
<td>4</td>
<td>We had a Board member ask for us to institute some sort of customer satisfaction score for the dashboard that goes to the board on a monthly basis</td>
</tr>
<tr>
<td>DEM: External Pressure</td>
<td>DEM-E</td>
<td>7</td>
<td>I just had had to do a report for our donors for youth programs . . . and they just asked me for stats. Like how many people came to our programs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resources/Supports</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;S: Human Resources</td>
<td>R&amp;S-HR</td>
<td>25</td>
<td>There’s Keri who is the head and then there are two interpretive planners and two assistant interpretive planners</td>
</tr>
<tr>
<td>R&amp;S: Time</td>
<td>R&amp;S-T</td>
<td>3</td>
<td>All of us have in our job description that about 20% of our time is supposed to be devoted to visitor research</td>
</tr>
<tr>
<td>R&amp;S: Financial Resources</td>
<td>R&amp;S-$</td>
<td>1</td>
<td>Sometimes we’ll get a program that comes with funding and then we might be able to evaluate it but since I’ve been in the role we haven’t actually had the chance to do any of that</td>
</tr>
<tr>
<td>R&amp;S: Other</td>
<td>R&amp;S-O</td>
<td>4</td>
<td>The other thing that I did that I think really made the biggest difference was I put together a visitor research committee and marketing was on it and membership was on it and visitor services was on it and education was on it and we all started to talk about our data across the whole institution</td>
</tr>
<tr>
<td>R&amp;S: Technology</td>
<td>R&amp;S-T</td>
<td>7</td>
<td>So, for the gallery guide program we have a system called the volunteer information system where we can enter what we call activities</td>
</tr>
</tbody>
</table>

- **CUL**: CULTURE
- **DEM**: DEMAND
- **R&S**: RESOURCES/SUPPORTS
<table>
<thead>
<tr>
<th>Capacity</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP: Knowledge and Skills</td>
<td>CAP-K&amp;S</td>
<td>9</td>
<td>I did an undergraduate degree in psychology which is why I ended up in visitor research because I did a lot of statistics and methodology and then I did Museum Studies at the University of Toronto</td>
</tr>
<tr>
<td>CAP: Instrumental Use</td>
<td>CAP-I</td>
<td>17</td>
<td>We’ll be able to determine whether we can continue with the program in its current structure because it is fairly costly for us</td>
</tr>
<tr>
<td>CAP: Conceptual Use</td>
<td>CAP-C</td>
<td>19</td>
<td>What we are able to do is, based on your postal code, identify “are you an urban digerati for example?” These are profiles of people with similar interests and it helps us to know who is coming</td>
</tr>
<tr>
<td>CAP: Persuasive Use</td>
<td>CAP-Per</td>
<td>3</td>
<td>It’s also a bit of proving cause it’s new to this kind of world— the use of data—so for me it’s also trying to advocate the value of why good data is important</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activities</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT: Type of Evaluation</td>
<td>ACT-TOE</td>
<td>5</td>
<td>We try to get them to do different types of evaluations so we might have one group pick a front-end evaluation, another a remedial formative evaluation, and another group a summative evaluation</td>
</tr>
<tr>
<td>ACT: What Gets Evaluated</td>
<td>ACT-What</td>
<td>19</td>
<td>our resources go towards evaluating the success of our temporary ticketed exhibitions</td>
</tr>
<tr>
<td>ACT: Methods</td>
<td>ACT-M</td>
<td>41</td>
<td>we do a monthly exit survey which is a pretty traditional six question questionnaire which is administered by someone who works for me</td>
</tr>
<tr>
<td>ACT: Type of Information Collected</td>
<td>ACT-TOI</td>
<td>34</td>
<td>For courses because they’re revenue driven we have budgetary targets that we are striving for paired with a set number of registrations. So we measure those programs on successful retention and also just uptake</td>
</tr>
<tr>
<td>ACT: Who Was Involved</td>
<td>ACT-Who</td>
<td>11</td>
<td>The key performance indicators research project included 3 staff and one volunteer all working very closely together</td>
</tr>
</tbody>
</table>
Table Continued

<table>
<thead>
<tr>
<th>Activities</th>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT: Satisfaction</td>
<td>ACT-S</td>
<td>4</td>
<td>[Program evaluation] is certainly something we’re trying to do and we have done. It’s not like we don’t do it. We just don’t do it systematically and we don’t do it as well as we should</td>
</tr>
</tbody>
</table>

**ECB**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECB-I</td>
<td>4</td>
<td>Some of those tactics I learned from my colleagues by going to conferences and hearing about best practices and things that worked</td>
</tr>
<tr>
<td>ECB-D</td>
<td>2</td>
<td>I have taken a lot of certification through the Canadian Marketing Association for direct marketing and customer insights and analytics</td>
</tr>
</tbody>
</table>

**Challenges**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-R&amp;S</td>
<td>14</td>
<td>I only have a small budget and not a lot of time</td>
</tr>
<tr>
<td>CH-ID/C</td>
<td>3</td>
<td>When I was brought here it was to integrate visitor voices into exhibition or experience design. It wasn’t to evaluate whether or not our educational programs are really making a difference</td>
</tr>
<tr>
<td>CH-Prac</td>
<td>6</td>
<td>We thought we’d do some on site interviews with participants and the strategy we used didn’t work at all— no one participating wanted to talk to us on the night of</td>
</tr>
<tr>
<td>CH-Ed</td>
<td>5</td>
<td>Our resources go towards our big exhibitions because they bring in the most people. A program only serves fifty people and an exhibition serves 300 000, so that’s just the dirty side of it</td>
</tr>
<tr>
<td>CH-O</td>
<td>12</td>
<td>Hiring external evaluators is more difficult in a unionized environment so that was not possible here</td>
</tr>
</tbody>
</table>

**Needs**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Frequency</th>
<th>In-Text Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-R&amp;S</td>
<td>9</td>
<td>In my ideal world, I would have a role that was dedicated to program evaluation and if I had someone like that I would have the ability to evaluate some of our key programs</td>
</tr>
<tr>
<td>Needs</td>
<td>Abbreviation</td>
<td>Frequency</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>N: Internal Demand for and Organizational Commitment to Institutionalizing Evaluation</td>
<td>N-ID/OC</td>
<td>3</td>
</tr>
<tr>
<td>N: Training in Evaluation</td>
<td>N-T</td>
<td>2</td>
</tr>
<tr>
<td>N: Other</td>
<td>N-O</td>
<td>2</td>
</tr>
<tr>
<td>Reactivity/Opportunity</td>
<td>Abbreviation</td>
<td>Frequency</td>
</tr>
<tr>
<td>Reactivity</td>
<td>REAC</td>
<td>1</td>
</tr>
<tr>
<td>Opportunities</td>
<td>OPP</td>
<td>2</td>
</tr>
</tbody>
</table>
### Appendix L

**Sample of Documents and Documentary Data Analyzed Per Museum During Phase Two**

**The Montreal Museum of Fine Arts**

<table>
<thead>
<tr>
<th>Document Selected</th>
<th>Document Type</th>
<th>Documentary Data Analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>First art hive in a museum at the Montreal Museum of Fine Arts (Art Hives, 2017)</em></td>
<td>Press release</td>
<td>Contextual data on a new facility for multidisciplinary creative projects developed in collaboration with Concordia University’s department of Creative Arts Therapies</td>
</tr>
<tr>
<td><em>AAMD: Next practices in partnerships (Association of American Museum Directors, 2017)</em></td>
<td>Compendium of 95 submissions from AAMD’s membership exploring the many ways art museums work with different organizations</td>
<td>Contextual data on the MMFA’s partnership programming for underserved and vulnerable populations</td>
</tr>
<tr>
<td><em>Body talk: Examining a collaborative multiple-visit program for visitors with eating disorders (Baddeley, Evans, Lajeunesse, &amp; Legari, 2017)</em></td>
<td>Peer-reviewed article</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td><em>Preliminary evaluation of the Arts-Based Rehabilitation Group Program for youth with psychiatric disorders: Espace transition au musée (De Broux-Leduc et al., 2017a)</em></td>
<td>PowerPoint of paper presentation</td>
<td>Pilot evaluation objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td>Promoting recovery in youths with mental health disorders through an innovative arts-based group program (De Broux-Leduc et al., 2017b)</td>
<td>Poster presentation</td>
<td>Pilot evaluation objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td><a href="http://www.mbam.qc.ca/en/">http://www.mbam.qc.ca/en/</a></td>
<td>Website</td>
<td>Contextual data on the MMFA’s mission and programs</td>
</tr>
<tr>
<td><em>Rethinking docent training at the Montreal Museum of Fine Arts: A pilot project (Lachapelle, Keenlyside, &amp; Dousenard, 2016)</em></td>
<td>Peer-reviewed article</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td><em>Michal and Renata Hornstein Pavilion for Peace (MMFA, 2016)</em></td>
<td>Souvenir publication</td>
<td>Contextual data on the Michel de la Chenelière International Atelier for Education and Art Therapy and the manifesto for a humanist museum written by the Director General</td>
</tr>
</tbody>
</table>
### Table Continued

<table>
<thead>
<tr>
<th>Document Selected</th>
<th>Document Type</th>
<th>Documentary Data Analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2016-2017 annual report</strong> (Montreal Museum of Fine Arts, 2017a)</td>
<td>Annual report</td>
<td>Contextual data on the MMFA’s education and wellness programs and facilities</td>
</tr>
<tr>
<td><strong>Financial statements of the museum and museum foundation</strong> (Montreal Museum of Fine Arts 2017b)</td>
<td>Financial statements</td>
<td>Contextual data on the financial position of the MMFA and the results of its operations and cash flows</td>
</tr>
<tr>
<td><strong>A new art and health advisory committee for the Montreal Museum of Fine Arts</strong> (Montreal Museum of Fine Arts, 2017c)</td>
<td>Press release</td>
<td>Contextual data on the Art and Health Advisory Committee of the MMFA</td>
</tr>
<tr>
<td><strong>Clinical research unique in Canada on the effects of art on senior citizens</strong> (Med e-news, 2019)</td>
<td>Press release</td>
<td>Contextual data on clinical research program designed to evaluate the effects of art on people aged 65 and older</td>
</tr>
<tr>
<td><strong>The 30th edition of Entretiens Jacques Cartier finishes at the Montreal Museum of Fine Arts in the company of 200 arts and science professionals</strong> (MMFA, 2017d)</td>
<td>Press release</td>
<td>Contextual data on the Art and Health Advisory Committee of the MMFA</td>
</tr>
<tr>
<td><strong>Art therapy and wellness at the Montreal Museum of Fine Arts</strong> (MMFA, 2019j)</td>
<td>Brochure</td>
<td>Contextual data on the MMFA’s art therapy and wellness programs and research projects</td>
</tr>
<tr>
<td><strong>Day to celebrate Autism</strong> (MMFA, 2019l)</td>
<td>Press release</td>
<td>Contextual data on the <em>Art of Being Unique</em> research project headed by neurobiologist Bruno Wicker</td>
</tr>
<tr>
<td><strong>Recherches globales</strong></td>
<td>Internal Excel document</td>
<td>List of past, present, and future research and evaluation projects</td>
</tr>
<tr>
<td><strong>An adjunctive, museum-based art therapy experience in the treatment of women with severe eating disorders</strong> (Thaler et al., 2017)</td>
<td>Peer-reviewed article</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
</tbody>
</table>

### The Art Gallery of Ontario

<table>
<thead>
<tr>
<th>Document Selected</th>
<th>Document Type</th>
<th>Documentary Data Analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AGO year in review 2016–2017</strong> (AGO, 2017a)</td>
<td>Annual report</td>
<td>Contextual data on the AGO’s education and public programs and facilities</td>
</tr>
<tr>
<td><strong>AGO Audited Financial Statements 2016-2017</strong> (AGO, 2017b)</td>
<td>Financial statements</td>
<td>Contextual data on the financial position of the AGO and the results of its operations and cash flows</td>
</tr>
<tr>
<td>Document Selected</td>
<td>Document Type</td>
<td>Documentary Data Analyzed</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><em>Georgia O’Keeffe, April 2017</em></td>
<td>Internal exhibition evaluation report</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td><em>Guillermo del Torro: At home with monsters, March 2018</em></td>
<td>Internal exhibition evaluation report</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td><a href="http://ago.ca/">http://ago.ca/</a></td>
<td>Website</td>
<td>Contextual data on the AGO’s mission and programs</td>
</tr>
<tr>
<td><em>From consultation to collaboration: Mechanisms for integrating community voices into exhibition development (Koke &amp; Ryan, 2017)</em></td>
<td>Book chapter</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
<tr>
<td><em>Building diversity in museums (McIntyre &amp; Marcus Ware, 2009)</em></td>
<td>Peer-reviewed article</td>
<td>Contextual data on the AGO’s mission and programs</td>
</tr>
<tr>
<td><em>The idea of north: The paintings of Lauren Harris, November 2016</em></td>
<td>Internal exhibition evaluation report</td>
<td>Study objectives and methods of data collection, analysis, interpretation, and reporting</td>
</tr>
</tbody>
</table>