

Where's Betty? Integrating Gender-based Analysis Plus (GBA+) in Canada's Federal
Government Evaluation Function

Jane Whynot

A thesis submitted in partial fulfillment of the requirements for the
Doctorate in Philosophy, Feminist and Gender Studies

Institute of Feminist and Gender Studies
Faculty of Social Sciences
University of Ottawa

© Jane Whynot, Ottawa, Canada, 2021

Table of Contents

Table of Figures and Tables

Figure 1 Timeline of Betty Crocker achievements	vi
Figure 2 The integration of Gender-based Analysis Plus (GBA+) across the federal government policy cycle	20
Figure 3 Generic theory of change (ToC) and associated dimensions	41
Figure 4 Generic COM-B theory of change (ToC).....	43
Table 1 Tailored theory of change (ToC) for integrating Gender-based Analysis Plus (GBA+) in the federal government evaluation function	60
Table 2 Critical resource review linking Gender-based Analysis Plus (GBA+) and evaluation	72
Table 3 Heads of Evaluation (HoE) survey outreach	83
Figure 5 Identified values of integrating Gender-based Analysis Plus (GBA+) in evaluation	87
Figure 6 Identified benefits of Integrating Gender-based Analysis Plus (GBA+) in evaluation..	88
Figure 7 Current and future projections of evaluation function intersectional analysis	123
Figure 8 Updated Theory of Change (ToC)	123

List of Abbreviations

ACOA	Atlantic Canada Opportunities Agency
ADM	Assistant Deputy Minister
AEA	American Evaluation Association
AG	Auditor General
AGE	Agenda for Gender Equality
AoM	Area of Management
AoM6	Area of Management Evaluation
AU	Australia
cs-QCA	Crisp Set Qualitative Comparative Analysis
CA	Capabilities Approach
CAE	Chief Audit Executive
CAS	Capacity Assessment Survey
CES	Canadian Evaluation Society
CESEF	Canadian Evaluation Society Educational Fund
CGC	Comptroller General of Canada
CIC	Department of Citizenship and Immigration
CIHR	Canadian Institutes of Health Research
CIRNA	Crown Indigenous Relations and Northern Affairs Canada
COM-B	Capacity Opportunity Motivation Behaviour
CRE	Culturally Response Evaluation
CSPS	Canada School of Public Service
DEC	Departmental Evaluation Committee
DEP	Departmental Evaluation Plan
DFO	Department of Fisheries and Oceans Canada
DG	Director General
DH	Deputy Heads
DM	Deputy Minister
DOF	Department of Finance
DP	Departmental Report
DPG	Designated Priority Groups
DRF	Departmental Results Framework
EA	Economic Affairs
EDI	Equity Diversity and Inclusion
EES	European Evaluation Association
ESDC	Economic and Social Development Canada
FEWO	Standing Committee on the Status of Women
fs-QCA	Fuzzy Set Qualitative Comparative Analysis
FTE	Full-time Equivalent
GA	Government Affairs
Gs&Cs	Grants and Contributions
GAC	Global Affairs Canada
GAD	Gender and Development
GBA	Gender-based Analysis
GBA/+	Gender-based Analysis and Gender-based Analysis Plus

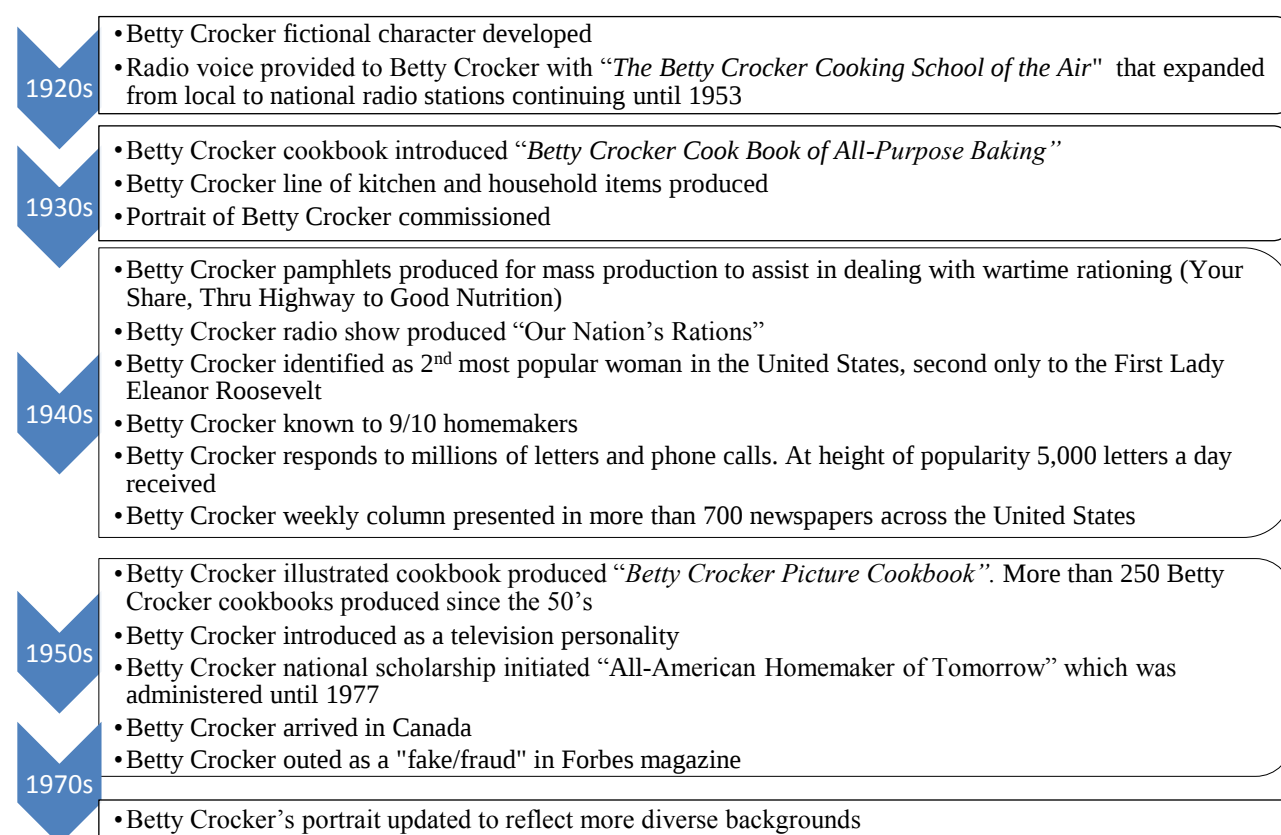
GBA+	Gender-based Analysis Plus
GC	Gender Champions
GEDS	Government Employee Directory Services
GEEW	Gender Equality and Empowerment of Women
GFP	Gender Focal Points
GM	Gender Mainstreaming
GRF	Gender Results Framework
HC	Health Canada
HoA	Head of Agency
HoE	Heads of Evaluation
HoPM	Head of Performance Measurement
HR	Human Resources
HR	Human Resources
IA	International Affairs
IFGS	Institute of Feminist and Gender Studies
INAC	Indian and Northern Affairs
IRCC	Immigration, Refugees and Citizenship Canada
ISC	Indigenous Services Canada
IT	Information Technology
KASA	Knowledge, aspirations, skills and attitudes, acquisition
L&L	Lunch and Learn for Evaluators
LDA	Large Department or Agency
MAF	Management Accountability Framework
MC	Memorandum to Cabinet
MRAP	Management Response and Action Plan
NCC-CES	National Capital Chapter-Canadian Evaluation Society
NCR	National Capital Region
NZ	New Zealand
OAG	Office of the Auditor General
OCAP	Ownership, Control, Access and Possession
OGD	Other Government Departments
OGS	Ontario Graduate Scholarships
PACP	Standing Committee on Public Accounts
PCH	Department of Canadian Heritage
PCO	Privy Council Office
PHAC	Public Health Agency of Canada
PIP	Performance Information Profile
PM	Prime Minister
PMEC	Performance Measurement and Evaluation Committee
PMO	Prime Minister's Office
PPX	Planning and Performance Exchange
PWCSC	Public Works and Consulting Services Canada
QCA	Qualitative Comparative Analysis
ROI	Return on Investment
SMV	Set Membership Value
SA	Social Affairs

SDA	Small Department or Agency
SDG	Sustainable Development Goal
StatsCan	Statistics Canada
SWAP	System Wide Action Plan
SWC	Status of Women Canada
TB	Treasury Board
TBE	Theory-based Evaluation
TB Subs	Treasury Board Submissions
TBS	Treasury Board Secretariat
TIG	Topical Interest Group
ToC	Theory of Change
TRC	Truth and Reconciliation Commission of Canada
Ts&Cs	Terms and Conditions
UK	United Kingdom
UN	United Nations
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
UN-SWAP	United Nations System Wide Action Plan on Gender Equality and the Empowerment of Women
WAD	Women and Development
WAGE	Women and Gender Equality
WID	Women in Development

Betty Crocker

Betty Crocker was a fictional domestic creation providing baking, cooking and home guidance to homemakers across the United States from the 1920s onward. The purpose driving the creation of this domestic goddess was to provide individualized, feminized responses to homemaking challenges. Despite Betty Crocker's outing as a fake/fraud by Forbes Magazine, at the height of her popularity more than 5,000 fan letters were received daily. This amounted to more than 1 million individualized responses via radio shows, newspaper advice columns and television appearances during the 40's and 50's. My grandmother's adoration of Betty Crocker is reflected in chapter headings that adopt the advice seeking format to frame content. A brief timeline of Betty Crocker's achievements is presented below in Figure 1.

Figure 1 Timeline of Betty Crocker Achievements



Abstract

This doctoral research by Jane Whynot explores how Gender-based Analysis/Plus (GBA/+), has been integrated into the Canadian federal government's evaluation function. This research has been supervised by Dr. Caroline Andrew at the University of Ottawa. Efforts presented in this dissertation have been profoundly influenced by political leadership, and their efforts to reprioritize gender equality and other elements of diversity across the bureaucracy. Presumably, employing this mechanism across the policy cycle will provide information to decision-makers that will ultimately impact diversity, equity, and inclusion (EDI) outcomes. Canada is unique amongst governments worldwide in its efforts to so comprehensively consider personal identity elements in policy discourse. Sadly, however, the federal government's evaluation function is lagging. Both domestic and international cues prompting the integration of gender and other elements of diversity into evaluation efforts have fallen unheard despite official domestic commitments formalizing these linkages.

Theories of intersectionality, and how intersectional analysis is undertaken are at the core of this research. The "+" in GBA was intended to represent all elements of personal identity beyond gender such as but not limited to race, language, sexuality, geographic location, disability, and Aboriginal status. The federal government's lead agency for gender equality does not explicitly reference details on any of these components in their publicly facing materials. These elements are embedded in this research's conceptual framework and subsequent lines of evidence. Michie, VanStalen, and West's (2012) theory of behavioural change provides the foundational structure for this research. These authors suggest that (c)apacity change only occurs when (m)otivations, (o)opportunities, and c(apabilities) are simultaneously present. Mayne (2016, 2017, 2018) was responsible for integrating this model of behavioural change into

program theory; it is this COM-B theory of change (ToC) that has been used as the conceptual framework for this research. In Mayne's COM-B ToC, capacity change outcomes are situated mid-results chain and are accompanied by both precedent and antecedent outcomes. A tailored ToC for integrating GBA+ in the federal government's evaluation function is presented in the thesis that aligns with the changing state of GBA+ implementation across the policy cycle, honing in on the federal government evaluation function. This research was initiated before the specific articulation of GBA+ as a federal government priority in the *Policy on Results* (Treasury Board Secretariat, 2016); the conceptual framework serves to capture the dynamic nature of the state of GBA/+ implementation across the federal government's policy cycle.

This research adopted a mixed-methods approach involving qualitative and quantitative lines of evidence as recognized by Greene (2007). Supporting lines of evidence included:

- A survey was administered to the Heads of Evaluation (HoE) who are senior level decision-makers responsible for the evaluation function within each federal government organization. Survey questions posed addressed the entire tailored ToC. Responses were intentionally designed to solicit binary responses;
- A critical review of all publicly accessible federal government produced resources exploring the linkage between gender mainstreaming, GBA and GBA+ and evaluation;
- Key informant interviews comprised of representatives from various federal government organizations including leads from central agencies and those responsible for gender mainstreaming, academics with expertise in feminist evaluation, Equity Diversity and Inclusion (EDI), and
- Case studies of federal government evaluation functions exploring the integration of GBA+ in evaluation efforts (n=6). Each case study focused on a different element of

GBA+ evaluation integration. Focus areas of each case study included: 1 organization focused on building evaluation function capacity to integrate GBA+, 1 organization focused on building capacity to integrate GBA+ amongst the evaluation and performance measurement and/or results unit, and 4 organizations focused on a single program. Each case study included the following lines of evidence: a literature review, a review of evaluations for GBA+ since 2016, a review of organizational documentation, key informant interviews, and for 3 case studies, a combination of focus groups/information exchange sessions.

The COM-B tailored ToC for integrating GBA/+ in the federal government's evaluation function encompasses a causal outcome pathway represented by a results chain. These pathways map a series of outcomes extending from activities and outputs, stakeholder reach/reaction, capacity change, behaviour change, direct benefits, and improved well-being. Each of these outcomes has multiple accompanying assumptions. Data from the tailored ToC is presented across thesis chapters. Despite widespread recognition of nascent GBA+ policy requirements and the federal government's GBA+ tool, little GBA+ instrumental evaluation use has resulted. A wide interval exists however between awareness and use. Nestled between building GBA+ capacity and instrumental use within evaluation functions, are multiple examples of GBA+ conceptual and process use. Many organizations are aware of the federal government's GBA+ tool, but there is inconsistent uptake of related training opportunities to build this capacity. Many federal organizations have undertaken only comparative or additive intersectional analysis and indicated limited intention to progress to interactional or truly intersectional integration analysis. To some extent, evaluation functions must navigate their course in this uncharted territory. Federal government organizations identifying as having established some capacity for GBA+

integration have developed unique GBA+ evaluation tools specific to their organizations in the absence of central agency guidance materials and a lack of evaluation function engagement of Gender Champions (GC) and Gender Focal Points (GFP).

This research provided the first examination within the governmental contexts of efforts to integrate intersectional research within evaluation functions. While providing answers to many questions, many more arise – consequently providing the foundation for a potential future research program.

Acknowledgements

As a mature graduate student, I had no idea what lay ahead when I naively entered doctoral studies in 2014. I had no idea how many organizations and individuals would support me during this time; their unflagging assistance made the difference between completing my doctoral studies and finishing this research.

Organizations providing financial support were both academic and from my chosen discipline (evaluation). Gratitude is extended to the University of Ottawa; Ontario Graduate Scholarships (OGS) as well as the Planning and Performance Exchange (PPX) and the Canadian Evaluation Society Educational Fund (CESEF). Without this financial support, this research would not have been possible.

To the academics guiding my research, I also owe gratitude. My committee was comprised of a diverse set of individuals from all areas of academia ranging from governance and public administration, education, evaluation and feminist studies. Before applying to the Institute of Feminist and Gender Studies while searching for a supervisor, it was shared with me that Dr. Caroline Andrew would be a dream supervisor. It was also noted however that chances were slim that she would accept me. Caroline was indeed a dream supervisor who was always supportive, responsive to various inquiries in a timely fashion, and more importantly guided my research with a gentle hand. I appreciate being one of her last students more than she would ever realize. Dr. Robert Shepherd from Carleton University joined my committee as my evaluation sounding board and his critical view of the Canadian government's evaluation function was always appreciated. Dr. Angel Foster with her background in women's reproductive health was appreciated for her logical insights and willingness to engage from afar. Finally, Dr. Ruth Kane's contributions as a late committee edition, thank you for your willingness to engage despite your

unfamiliarity with the context of my research. I would also be remiss if I did not mention my friend and colleague Dr. Isabelle Bourgeois, now at the University of Ottawa. Her words of advice during some of the more troubling moments were what saw me through this trying academic period. I have no doubt that without these individuals, I would have achieved far less professional and academic success without their contributions.

To the friends and family across Canada who vicariously supported these studies I cannot express how thankful I am for your contributions. It may seem insignificant to have your stove replaced when it unexpectedly explodes, gift certificates for eyebrow threading appear, random books appearing by mail; your dog cared for during the writing months, the unsolicited meals that would appear at my door, the late night borrowing of pools when the dog had been skunked, critical feedback on dissertation drafts, or the seemingly random phone calls. Each gesture was interpreted as a demonstration of the love and support offered by each of you and were noted and appreciated. These individuals include Jaki Freeman, Tracey Hubley, Darryl Stewart, James Coyle, Alan Fong, Renée-Pierre Charron, Lisa and Donal O'Reilly, and Meghan Beaton.

To the members of the evaluation community in Canada and further abroad, thank you. There are far too many to name to name; from the professional association boards that I serve with; the journal that I volunteer with; to those I have mentored and have been mentored by; to those individuals in case study organizations onboarded largely because longstanding personal and professional relationships. Your support throughout this process has been boundless. This is a community that I feel passionately about and have invested countless hours in because I believe that we can make a difference for all individuals. I owe a special thanks to Steve Montague who was one of my first instructors in program and policy evaluation. He fostered my

professional development over the years, and I count myself fortunate enough to have been present for many shared moments of insightful brilliance.

To my family members, you have my eternal gratitude. Your words of advice and financial support made a difference. I cannot count all the ways possible in which your encouragement made it possible to remain dedicated to this research for such a sustained duration. Mom and Scott, thank you for your emotion and financial assistance. Importantly, I also need to thank my grandmother Winnifred Jackson who passed away in 2006. It is my reflections on our earliest times together baking in her kitchen that structured the format of this dissertation. Much like Betty Crocker herself, she was able to offer critical insights into my life and those around her with a few simple words.

Chapter 1 - Dear Betty: My cake is flat. What do I do?

This chapter describes the history and background of gender mainstreaming in Canada and its progression to gender-based analysis (GBA). When this research was initiated, gender-based analysis plus (GBA+) had not yet been formally introduced within the federal government environment; the transition is addressed in later chapters of this dissertation. An introduction to the Canadian federal government's evaluation function is provided as well as a description of how linkages between gender-based analysis and the evaluation function have been formalized based on approaches employed by the international evaluation community. In addition, a discussion of why integrating GBA in the Canadian federal government evaluation function is important is tabled.

Gender mainstreaming in Canada

The process of embedding personal identity factors, including gender, into policy discourse is broadly referred to as gender mainstreaming. Although gender mainstreaming definitions vary, consensus exists on two dimensions that include the prioritization and achievement of gender equality objectives to address gaps between the current status quo and future visions of what various equalities might look like. Gender mainstreaming is a strategy to close gaps between various identity expressions that both drive and result in inequalities. To quote the scholarly literature, “[gender mainstreaming] is a policy about policy” (McNutt & Béland, 2004, p.469).

Gender mainstreaming gained traction in European countries prized for their innovative approaches for their centralization of gender in public policy discourse. These efforts were mirrored in the Canadian context. Canada historically has been internationally renowned for its support and promotion of not only gender but other axes of diversity within the public administration domain. Significant federal initiatives investigating various intersections of

diversity have been instigated, including Royal Commissions. Royal Commissions respond to publically identified social or economic controversies; paving future government policy directions (Lockwood, 1967) while simultaneously mollifying politically influential public stakeholders. The focus of these diversity initiatives have independently examined many intersections of diversity that include pluralism (culture and language, multiculturalism (culture, race, immigrant status, and religion), the rights of Aboriginal and First Nations groups, and include the Status of Women in Canada.

In 1967, the Royal Commission on the Status of Women in Canada initiated a consultation process that solicited input from both the public and private sectors resulting in the tabling of 167 recommendations in the House of Commons addressing multiple axis expressions of diversity. Recommendations were framed under the requirement for systemic responses to address inequities through the implementation of *policy* (emphasis added) (Parliament of Canada, 1970). The Commission report referenced that “we cannot ignore the larger issues of customs and beliefs which have given rise to discrimination against women in the past, which continue today, and which may impose new forms of injustice in the future” (p.387). The majority of the recommendations were at least partially implemented by the 1980’s; however, two critical failings emerged. These included an almost total disregard of women’s experiences of poverty, and a failure to address gender-based violence (Morris, 2006). Differences in key indicators of gender equality representing inequalities experienced by women are pervasive and continue through to the present time (Status of Women Canada, 2015). Like the Royal Commission on the Status of Women in Canada report recommendations, implementation of other recommendations of Royal Commissions related to diversity have also been met with varying degrees of success in

terms of their subsequent entrenchment in federal government programs and policies (Doerr, 2006; Laing, 2013; Morris 2006).

The Council of Europe (2004) identified seven preconditions necessary to support successful implementation of gender mainstreaming initiatives that include: political will; specific gender equality policy/ies; the systematic collection of data demonstrating understanding of gender relations; a comprehensive stakeholder understanding of gender relations; gender expertise knowledge centres; necessary resources including expertise and funding; and finally, the participation of women in political and public life and in decision-making processes. Canada's level of success in meeting these preconditions is varied (Atkins, 2017) where noteworthy deficits are reflected in four of the seven pre-conditions across multiple federal government departments. Questions remain as to why Canada as a nation, and particularly at the government level, we are not mainstreaming gender better.

Two strategic approaches to gender mainstreaming have been summarized by McNutt & Béland (2004) as being either integrationist or transformative. Integrationist gender mainstreaming occurs when a gendered lens, much like rose-coloured glasses, are shaded onto existing organizations and their policies. The integrationist approach assumes that doing so will not impede gender equality advancement. Alternatively, transformative gender mainstreaming occurs when reconceptualizations regarding how gender equality might be achieved are employed. Verloo summarizes three necessary building blocks underpinning how strategies of gender mainstreaming are considered that include "regular policies are gendered, that regular policies are a major constitutional element in the construction of gendered social institutions, and that gendered social institutions are an important component in the continuous reconstruction of gender inequality" (2001, p.3). Verloo's building block assumptions align with the interlocking

systems of oppression that can be traced back to Black feminism efforts of the 70's (Carastathis, 2016), and importantly, are a key dimension of what is currently embraced as intersectionality. Cumulatively, these building blocks encompass a systems' perspective which Rees (2000) positions as having the potential to affect change to lessen inequalities, particularly concerning gender mainstreaming. The emphasis on the latter two building blocks of gender mainstreaming was revolutionary (Verloo, p.8) because of their potential transformative nature. Either integrationist or transformative gender mainstreaming approaches have the potential to both interrupt the perpetuation of gender blind, or gender-neutral assumptions contributing to experienced inequalities, and to advance the current status quo assuming that action will be undertaken to change the current status quo.

It is important to note that various actors engaged in gender mainstreaming represent various layers of a system which are discussed in greater detail in the following chapters. This differs from previous gender equality efforts advocating for equal treatment between genders, and specific equality policies which situates different stakeholders such as legislators, and gender equality organizations as responsible for the implementation of various policies. Consequently, these actors ultimately shoulder significantly more of the change burden than other policy stakeholders. Systems approaches to gender mainstream allocate equal responsibility to all actors involved in the creation and implementation of policy. The assumption of diffused responsibility within this gender mainstreaming approach relies on bridge framing as forwarded by (Snow & Benford, 1986) in which boundary positioning widens so that broader meanings aligned with political discourse resonate with the newly broadened audience. In this expansion, messaging is key in positioning a shared value system, goals, and objectives.

General critiques of gender mainstreaming in Canada include institutional inertia, poor resource commitments, and weak administrative directives (Hankivsky, 2013) as well as competing operationalization priorities, poor integration processes, and weak accountability mechanisms (Meir & Lombardo, 2013). These factors have all contributed to weak, ineffective, and inconsistent gender mainstreaming in the federal government. Gender mainstreaming's action onus problematizes public decision making itself (Paterson, 2010) in presupposing that supplying sufficient information about differential outcomes experienced by diverse groups to enact changes that will lead to more informed decision-making which in turn, will lead to improved gender equality outcomes. Recent scholarly literature highlights the critical flaw in this conceptualization of gender mainstreaming in pointing out with unfailing certainty that posits, even with access to differential outcome information, decision-makers will act on this information.

From gender mainstreaming to Gender-based Analysis (GBA)

Since the 4th World Conference on Women in Beijing in 1995, more focused gender equality commitments were undertaken by the Canadian federal government (Paterson, 2010; Verloo, 2001; Scala & Paterson; 2017). As signatories to the Beijing Declaration for Action, the government committed itself to the “effective integration of a gender perspective throughout their operations, policies, planning, and decision making” (Parliament of Canada 2005, p.2). Subsequently, to honour its commitment Canada had initiated two five-year action plans that included the Federal Plan on Gender Equality (1995-2000), and the Agenda for Gender Equality (2000-2005). Gender-based analysis was featured predominantly in the two Status of Women of Canada gender equality framework documents to meet the commitments noted above.

Unfortunately, these commitments were not widely communicated, nor recognized by a significant proportion of federal government organizations.

A Standing Committee on the Status of Women (FEWO) was struck in 2004 as a result of recommendations made by the Royal Commission on the Status of Women in Canada. The committee's mandate includes "the broad authority to study the policies, programs, expenditures (budgetary estimates) and legislation of departments and agencies, including Status of Women Canada, that conduct work related to the status of women" (Parliament of Canada, n.d.). Throughout its existence, FEWO has extensively advocated, studied, and reported on GBA/+. Gender-based analysis is one of the primary mechanisms supporting reporting on gender equality's progress (Parliament of Canada, n.d.); evaluation is another (UNWomen, 1995). The Expert Advisory Committee on Accountability Mechanisms for Gender Equality identified the necessity of accountability mechanisms, including gender-based analysis (Parliament of Canada, 2005) for reporting on gender equality. This committee suggested that the degree to which gender-based analysis was implemented within federal government departments required further examination which eventually led to GBA examinations by the Office of the Auditor General (OAG) in 2009, and 2015.

While often referenced synonymously, important distinctions between gender mainstreaming and gender-based analysis (GBA). The distinction between GBA and gender-based analysis plus (GBA+) is described later in this thesis as it is a key dimension of the implementation story. Where GBA/GBA+ is used in this document, its use is intentional. Supportive relationships are intended but gender mainstreaming has been described as a policy and GBA/+ as a mechanism or tool. In the Canadian context, GBA/+ has become a substitute for gender mainstreaming. Gender mainstreaming is proactive and the focus tends to highlight

policy, legislation, and programs (McNutt & Béland, 2004). The subject is intended to address “structural discrimination and women’s exclusion from policy-making processes” (p.469). Using definitions supplied by SWC/WAGE, applications of GBA/+ can be strictly interpreted as reactive with its analysis/emphasis on the achievement of differential policy outcomes. This should be clarified however as only the federal government’s performance measurement and evaluation functions report on *actual* outcomes achieved. The remaining policy cycle junctures assume a proactive or ex-ante perspective of how individuals with different elements of diversity and focus on potential differences that *may* be realized.

Gender-based analysis was originally articulated as an approach “to work toward greater equality between men and women, and not an end in itself” (Parliament of Canada, 2005, p.3). Gender-based analysis then targeted both women-specific policies and ensured that other policies or legislations that were not women-specific did not exacerbate existing equality gaps. Early tool drafts of guidance materials for GBA were much more explicit (Status of Women Canada, 1998) about the need to consider social-cultural factors, and underpinning values. They also highlighted the need for consultation with members from diverse groups as part of the policy development process. As the gender equality discourse evolves in Canada, it has been interesting to witness the gender-based analysis tools’ evolution as well. What was originally a hard copy guidance document has evolved into a series of general online training modules and a plethora of related materials demonstrating how to tailor for individual organizations. Deviating from original GBA guidance materials, one key step “evaluating the quality of the gender-based analysis” has been dropped. This will be discussed in greater detail across the chapters of this dissertation. The latest addition of the Canadian federal government’s approach to gender mainstreaming occurred with the addition of one small symbol – the “+”. This symbol broadened gender mainstreaming

from sex/gender to all potential elements of diversity including but not limited to socio-economic status, religion, Aboriginal and visible minority status, sexuality, etc. Public-facing guidance materials note that the addition of this symbol rendered the approach intersectional which may be one of its key challenges. These challenges are discussed in greater detail in the following chapters.

Specific critiques of Canada's approach to gender mainstreaming inclusive of GBA/+ have emerged:

- it represents a culture change (Lombardo & Mergeart, 2013) and necessarily involves the adoption of a complexity lens;
- an emphasis on gender and sex as primary axis identity variables that can, at times, lead to the exclusion of all other identity factors (Hankivsky & Cormier, 2010; Paterson, 2010; Scala & Paterson, 2017);
- the privileging of GBA expertise including those with lived experience directly relevant or involved in the differential outcomes discourse. Gender-based analysis expertise becomes “uninterrogated and unchallenged” (Paterson, 2010, p.406);
- the positioning of the lead agency responsible for gender equality within the bureaucracy (McNutt, 2010);
- the lack of an accountability mechanism reporting on the implementation of GBA (Parliament of Canada, 2005; Office of the Auditor General, 2009; Office of the Auditor General, 2015);
- the allocation of sufficient resources (including expertise, financial resources, time) to build capacity to support implementation (McNutt, 2010; Scala & Paterson, 2017).

These weaknesses have been considered in the adoption of a theoretical framework and are discussed more fulsomely in Chapter 2.

It has been with the recent election of Justin Trudeau in 2015 that Canada as a federal government has renewed its longstanding commitment to gender equality. With this change in federal leadership, gender equality has emerged in the policy cycle in new and innovative ways that include a renewed emphasis on GBA/GBA+ through improved delivery of the GBA+ tool, mandatory inclusion of GBA/+ information in Treasury Board Submissions (TB Subs) and Memoranda to Cabinets (MC), a gender-responsive budgeting process, as well as a specific articulation of GBA in the 2016 renewal of the Policy on Results (Treasury Board Secretariat, 2016).

The Canadian federal government's evaluation function

Worldwide, Canada's claim to a continuous federal government evaluation policy is unsurpassed (Muller-Clemm & Barnes, 1997; Segsworth, 2005). Evaluation policy framework discussions were initiated in the late 60s by Treasury Board Secretariat (Muller-Clemme & Barnes, 1997, p.54) which, eight years later, subsequently resulted in the tabling of the first official evaluation policy known as 1977-47. The government's current definition of evaluation can be traced to these early efforts. Evaluation is defined as "the systemic collection and analysis on the outcomes of programs to make judgments about their relevance, performance and alternative ways to deliver them (Treasury Board Secretariat, 2016). Programs are interpreted in various ways and include programs, processes, initiatives, and as of 2009, policies (Treasury Board Secretariat, 2009). Objectivity featured significantly in this early evaluation policy (Muller-Clemme & Barnes, 1997) and was established via three criteria that included appropriate terms of reference, the independence of evaluators, and appropriate reporting levels. Also

highlighted in the policy circular, was the complementary relationship between evaluation and performance measurement.

This first policy followed a wave of financial accountability measures instigated by Bill C-20 in which greater management practices and controls (Treasury Board Secretariat, 2004) over the delivery and achievement of economy, efficiency, and effectiveness were demanded. National public administration discourse under the financial accountability umbrella realigned an emphasis from the resource-activity-output continuum to a focus on the outcomes produced as a result of government interventions. Key to the success of these financial accountability measures was the “information base on which such decision are made. Officials and ministers must have reliable, relevant and objective information available if real improvement in the decision-making process are to be achieved. Program evaluation is considered to be one important source of this information” (Treasury Board Secretariat, 2004, p.2). Program evaluation efforts in Canada were mirrored internationally; Treasury Board Secretariat citing the Organisation for Economic Cooperation and Development (OECD) noted that program evaluation is recognized because of its “importance in a results-oriented environment because it provides feedback on the efficiency, effectiveness, and performance of public policies and can be critical to policy improvement and innovation” (Organisation for Economic Cooperation and Development, 1999 in Treasury Board Secretariat, 2004).

More elaborate documents (Guide on the Program Evaluation Function and Principles for the Evaluation of Programs by Federal Departments and Agencies) were subsequently issued in following years identifying how program evaluation should be managed, and how evaluation should be undertaken under four key areas that included rationale, impact and effects, objectives achievement, and alternatives (Treasury Board Secretariat, 1981; Treasury Board Secretariat,

1984). Elements of these four components remain today. Across the two guidance documents, it is clear that the evaluation function was intended to support the management of financial accountabilities. During the 80s, approximately thirty departments across government were cumulatively on average undertaking one hundred evaluations each year (Rutman, 1986). This number fluctuated insignificantly over the next five years until 2004 in which a footnote in a Treasury Board Secretariat report on the quality of evaluations referenced approximately 250 evaluations being completed each year from 2002 onward (Treasury Board Secretariat, 2004, footnote1). Segsworth (2005) notes that the introduction of both a short-lived evaluation specific policy in 1991 and a replacement generic review policy encapsulating internal audit, performance monitoring, and several kinds of reviews in 1994 contributed to a decline in the production of evaluation reports in the 90s. These documents do not explicitly reference which stakeholders should be consulted during the course of an evaluation. Additionally, interpersonal relationships and skills are only acknowledged in so far as they predefine working relationships with senior management in the evaluation function (Treasury Board Secretariat, 1984, s.55).

Resulting from the 1986 government spending review (the Ministerial Task Force on Program Review [or the Nielson Task Force]), the potential of program evaluation to supply decision-makers with useful information regarding programs was deeply questioned. Subsequent examinations undertaken by those responsible for the government's evaluation function noted a misalignment between the criteria used to design government evaluations "useful". Two additional criteria were identified in defence of program evaluation-related findings of the Nielson Task Force that are particularly relevant to this research that included "making programs more responsive to client needs, ...[and], influencing policy decisions" (Rutman, 1987, p.21). Rutman's refutation of the Neilson Task Force's sweeping condemnation of the evaluation

function flagged future requirements for program evaluation to account for information availability at various levels including both program managers and decision-makers.

Synchronously, the requirement of performance measurement information to support evaluation efforts was also flagged as a future requirement. Subsequent studies of the relationship between evaluation and performance measurement noted damaging deficits and subsequently called for efforts to provide required data. As with any formal policy implementation process, many challenges ensued in subsequent years that included sufficient allocation of resources to meet evaluation demand, clarity for responsibilities regarding implementation within line organizations, and greater oversight and monitoring of related evaluation policies. Evaluation's contribution to decision-making can be summarized as having great potential, but ultimately falling short of this promise despite the orientation of additional evaluation policies towards outcomes in 2001, and 2009.

Introducing Gender-based Analysis (GBA) and evaluation

While gender equality's vision has remained steadfast in certain quarters within the federal government domain, its manifestation amongst the government's implementing organizations can be summarized as sporadic at best. Beijing's Declaration for Action identified three mechanisms as having the capacity for reporting on gendered impacts; these mechanisms include GBA, research, and evaluation (United Nations, 1995). In response to a Standing Committee on the Status of Women request, the Auditor General subsequently instigated an audit of GBA in federal government departments (OAG, 2009). Results were abysmal and concluded that even after 20 years of concerted efforts in building the capacity to implement GBA in various organizations across government, that implementation was uneven, and had demonstrated little impact on decision-making. The Parliamentary Standing Committee on Public Accounts (CPAC)

noted specifically that “the OAG had found departments provided limited information to the central agencies on the gender impacts of proposals. There was no reference to gender impacts in 15 of the 28 MCs [Memorandum to Cabinet] and in 8 of the 21 TB [Treasury Board] submissions that were examined. There was no indication why this information was not reported”. (Standing Committee on Public Accounts as cited in OAG, 2009, p.5). Critical flaws were also noted in departmental organizational GBA frameworks, evidence of GBA conducted by central agencies, a lack of monitoring and evaluation efforts related to the extent to which GBA had been undertaken. Stemming from the first Auditor General (AG) report, however, was the first-ever mandated collation of GBA/GBA+ and evaluation functions to include representations of diversity in performance stories. Efforts were initiated between Canada’s lead agency for gender equality and the federal government’s evaluation policy centre to produce guidance materials for federal government evaluators on how to integrate GBA in evaluation efforts. The government response tabled in Parliament in late 2010 recognized these efforts:

While the GBA e-Basics course is aimed towards analysts in general as an introductory course, the functional specific training is being developed to focus on the integration of GBA to specific and more complex or highly technical functions, such as evaluation and research, which require more specialized information. The goal is to make GBA relevant to different categories of professionals in the public service. (Government of Canada Response, 2010)

Although these originating efforts were scaled back with the lack of agreement between these organizations, it served as an important symbolic first step in formalizing the relationship. In 2015, the AG undertook a follow-up report on gender-based analysis which was tabled in early 2016. In this follow-up report, recommendations were made to undertake “concrete actions to identify and address barriers to preventing the systematic and rigorous conduct of GBA during the development, renewal, and assessments of policy, legislative and program initiatives to inform decision-makers” (Office of the Auditor General, 2015). It was further noted that “the

departments that had implemented a GBA framework did not always conduct complete analysis, and the quality of their analyses was not consistent” (Office of the Auditor General, 2015).

Quality of GBA included preliminary (and insufficiently) criteria that included: consideration of questions regarding gender and diversity, presentation of the implications, recommendations provided reflecting these considerations, and the provision of evidence regarding how gender impacts are aligned with other government priorities (Office of the Auditor General, 2009). The role of the evaluation function as potential producers of gender and diversity impact information was publicly recognized; central and gender equality agencies agreed to work together to produce related guidance materials for the federal evaluation function in response to the OAG recommendations. Additionally, Treasury Board Secretariat agreed to two interrelated actions that included: 1) the dissemination of these guidance materials related to the integration of GBA and the evaluation function; and 2) the examination of options to help departments and agencies consider government-wide policy considerations, such as GBA, when assessing the performance of federal programs (Office of the Auditor General, 2015). These concrete actions coupled with the change in federal leadership to a more liberal and self-identified “feminist” political leader coincided with several relevant results policy changes in 2016.

With epistemological shifts in political discourse accompanying political leadership transitions, for the first-ever since the issuance of the 1977-47 evaluation policy circular, GBA was specifically articulated as a federal government priority in the renewed suite of results policy tools in 2016. While caveated with “where relevant”, evaluators were policy-directed to consider GBA in both evaluation processes, and products. This represents a departure from other developed countries’ evaluation practices which were consistent with established Canadian practices until the release of the innovative 2016 *Policy on Results*. Accompanying this new

policy for evaluation, stronger linkages with other functional communities necessary to support GBA implementation such as performance measurement, gendered expertise, policy, and research were also established through the inclusion of relevant GBA examples in the *Interim Guidance on the Policy on Results* (TBS, 2017). The inclusion of GBA in these materials is somewhat ironic given that the earliest Comptroller-General of Canada (CGC) responsible for the evaluation function in its formative years had thematically noted across a review of several evaluations “...that services were being used disproportionately by those who were able, or who ought to be able to communicate with their government. The less educated, the less mobile, the less articulate were not being reached as much as might have been hoped” (Rayner, 1986, p.7).

Inarguably, this leads to questions regarding the efficient use of public resources and additionally leads to wondering about the unrealized potential of both the GBA and the evaluation functions. What would have occurred if a more somber reflection of the integration of these two functions been initiated much earlier? Despite the support provided by related policy frameworks, gender and other intersections of identity have been integrated into the evaluation practice dimensions in only a limited fashion. Pockets of evaluation practice reflecting an integration of dimensions of diversity are strong within federal government programming areas mirroring international commitments in areas such as development, health, and gender equality. Benefits derived from incorporating gender, and other elements of diversity into evaluation practice are numerous. Despite these benefits, as Bustelo (2016) notes government decision-makers have largely ignored their potential.

Cues from the international domain on integrating gender and human rights in evaluation

Integrating gender in evaluation involves considering not only gender and diversity-related outcomes of programs, but also the inclusion of related considerations in the evaluation design,

processes, products, and use (Lan, Dodd, Whynot & Skinner 2018). How gender is integrated into evaluation efforts varies; it is dependent upon ideological choices, methodological, political, and logistical considerations (Bamberger, Mabry & Khan, 2020). The adoption and implementation of gender across each of these evaluation dimensions pose unique challenges for evaluation stakeholders. The international evaluation community is driven largely by the family of United Nations (UN) organizations as well as World Bank affiliated organizations who have seriously refocused gender targeting both the evaluation policy and practice dimensions in the last twenty years. Motivations driving these efforts are partly attributable to the renewal of the sustainable development agenda in which gender equality and women's empowerment are standalone goals in the renewed development agenda, but have been embedded in several of the remaining 16 Sustainable Development Goals (SDGs) (Langford 2012; Fukuda-Parr, Yamin & Greenstein, 2014). Historically, multilateral organizations have focused their attentions on the metrics associated with women's economic empowerment (WorldBank 1994; United Nations Children's Fund, 2006) and have garnered severe critiques for this focus (Anderson & Langford, 2012; Kabeer, 1999; Perezniето & Taylor, 2014) versus embracing gender equality for its significance. Often governments refer to these organizations comprising the international evaluation community because of their long-standing experience in dealing with inequalities/discrimination, and their building of related capacities.

The United Nations System Wide Action Plan on Gender Equality and the Empowerment of Women (UN-SWAP) is the first system-wide accountability framework for monitoring gender mainstreaming across UN organizations. It is comprised of "15 performance indicators for tracking six main elements on gender mainstreaming: accountability, results-based management, oversight, human and financial resources, capacity, and knowledge exchange and networking

(Barnes & Bishop, 2016, p.4). Evaluation plays a critical role in UN-SWAP serving as an accountability mechanism to ensure that evaluation units are normalizing gender in evaluation. In 2016, the United Nations Evaluation Group (UNEG) initiated a review of UN-SWAP. The initial sample to be included in this review highlighted 378 evaluation reports from across 25 UN entities. This sample was reviewed and categorized as either ‘missing’ (11%), ‘meeting’ (37%), ‘approaching’ (33%), or ‘exceeding’ (19%) evaluation performance indicator (EPI) requirements related to SWAP. Distillation efforts resulted in 46 minimum and high-scoring evaluation reports included in exercise to assess trends, challenges, and good practices in integrating gender equality and human rights in evaluation efforts (Barnes & Bishop 2016). At a minimum in practice, directives have been provided that outline each UN entity should be ‘meeting’ EPI requirements. The application of quality criteria to the initial sample and subsequent reduction of evaluation reports for inclusion alludes to significant opportunities to improve the integration of gender equality and human rights in evaluation efforts within the UN system. While increasing attention and action are placed on the evaluation of gender mainstreaming in UN entities, only minimal attention is reflected in other sectors including the government.

Evaluation practitioner guidance issued for the international development community demonstrates a commitment (United Nations Children’s Fund, 2011; United Nations Evaluation Group 2015; UNWomen, 2015) about the varied ways in which gender equality and human rights may be incorporated in evaluation efforts. Historical deficits failing to capture both individual and systems-oriented thinking regarding social, cultural, political, and economic contexts have been addressed by situating the integration of gender in evaluation efforts by embedding systemic evaluation efforts with intersectional analysis in complex environments (Stephens, Lewis & Reddy, 2018). Additional guidance materials have been developed that

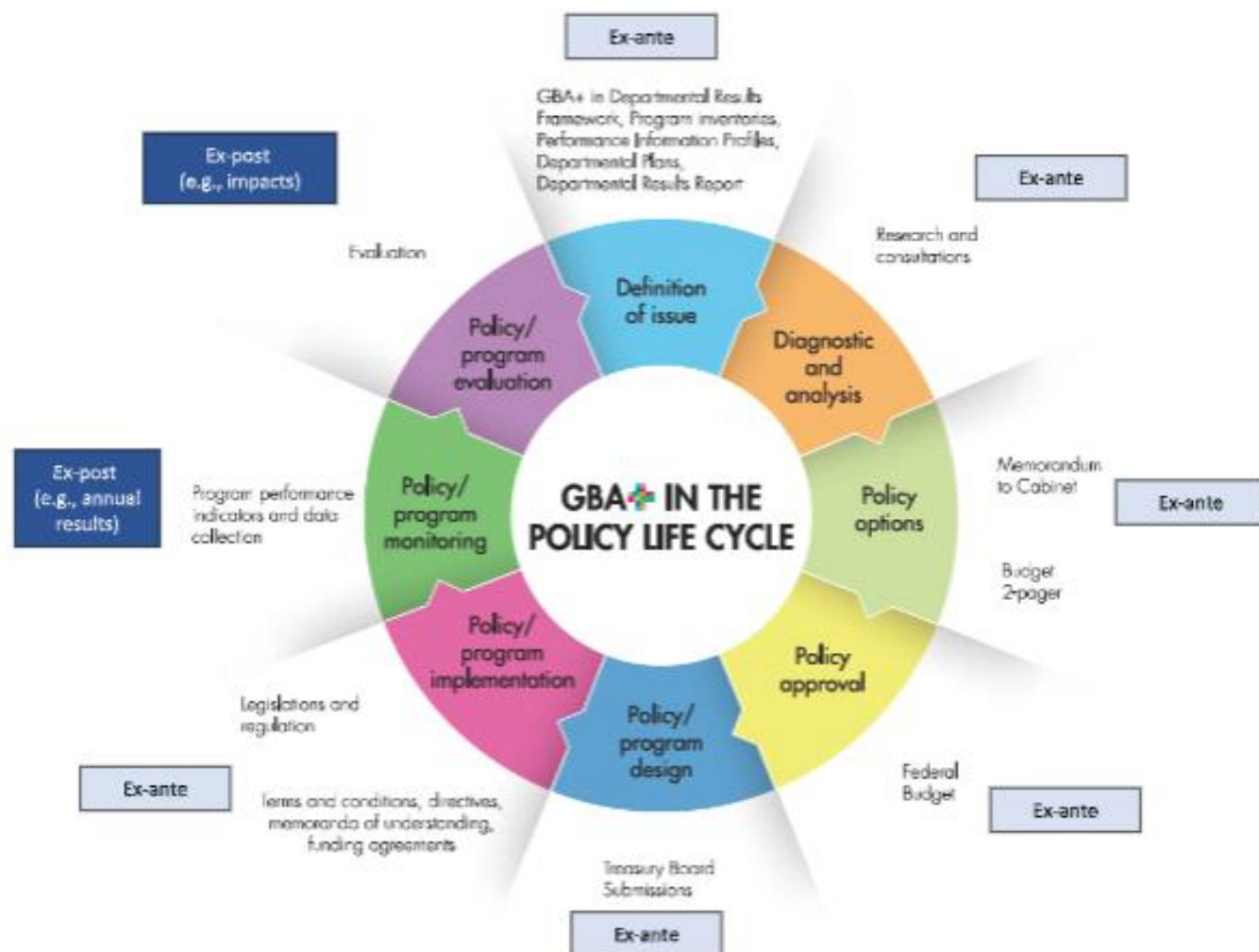
target the role of evaluation policies to support the integration of gender equality and social equity (Bamberger, Segone & Reddy, 2015). These materials identify eleven components necessary to integrate gender that include rationale, objectives/purposes, principles, responsibilities, coverage, methodology/techniques, ethics, dissemination, use, and reviews (Bamberger, Segone & Reddy, 2015).

The importance of integrating Gender-based Analysis Plus (GBA+) in evaluation

Finally, the axiological nexus of this research where GBA+ intersects with the federal government's evaluation function is arrived at. It is important to acknowledge that no other bureaucracies have approached gender mainstreaming as comprehensively across the policy cycle as Canada. This includes governments of the United Kingdom (UK), Australia (AU), New Zealand (NZ), and the Netherlands. While each of these jurisdictions has implemented gender mainstreaming across elements of the policy cycle, specific junctures are targeted and none as comprehensively as in Canada. Figure 1: The Integration of GBA+ Across Canada's Policy Cycle presented below, highlights where, and what policy requirements include a GBA+ element. When this research was initiated, only 1 GBA+ requirement was present; this requirement was the 2007 Treasury Board Submission GBA+ requirement (TB Sub). The implementation of other GBA+ requirements across the policy cycle followed rapidly, including the 2016 Memoranda to Cabinet (MC) GBA+ requirement as well as the 2017 federal budget GBA+ requirement. Additions of ex-ante, and ex-post labels reflect where GBA+ could be considered to be either predictive (results that could potentially be realized), or reflective (results actually achieved) respectively. Across the entire policy cycle, only the performance measurement and evaluation functions adopt an ex-post perspective. This means that these 2 functions are the only ones directly reporting the results achieved by federal government

spending. In the future, the federal government evaluation function could potentially be able to speak to what effect federal government spending has had on the lives of individuals with diverse identity elements should GBA+ be implemented as intended. For this reason, these functions are incredibly important. In the future, the performance measurement and evaluation functions should produce GBA+ data and outcome information that directly informs the front end of the policy cycle. At this point, ex-post GBA+ contributions cull together information from a variety of sources including academic offerings from within the country and further abroad, consultations with community service organizations, etc. to estimate potential results that may be achieved. Rarely are evaluations used to feed into these front-end policy cycle GBA+ requirements which is an inefficient use of resources. Both the federal government performance measurement and evaluation functions are referenced above, however, the focus of this research highlights the evaluation function.

Figure 2 The integration of Gender-based Analysis Plus (GBA) across the federal government policy cycle¹



This research explores in greater depth what the integration of GBA+ in the federal government evaluation function entails. Superficial complementarities exist between the two mechanisms including an emphasis on assessment competencies. Results from any assessment activities undertaken by either the GBA/+ and evaluation functions are intended to inform

¹ *This diagram has been updated by the researcher to include ex-post/ex-ante dimensions. The diagram was originally supplied by Treasury Board Secretariat during their consultation on the production of a central agency driven annual gender and diversity impact reporting exercise

decision-making. Studies of GBA/+ in Canada focus on knowledge production, particularly the supply of GBA/+ information to decision-makers (Paterson, 2010; Lombardo & Mergaert, 2013; Paterson & Scala, 2017). Evaluation theorists, however, have long hypothesized and demonstrated that information is merely one component of organizational decision-making (Weiss, 1998). Other similarities unite the 2 mechanisms. Given its feminist origins and potentially off-putting associations with feminist epistemologies, GBA+ could also be identified as political. Furthermore, by definition embedding feminist situated knowledge creation within the evaluation function and its ancestry embedded in neutrality could potentially create additional tension. Where these two mechanisms differ lies in the specificity of GBA/+’s attention to potential variances in impacts experienced by various program beneficiaries, and the vastness of any noted differential.

The core hypothesis guiding this research is that the integration of these two mechanisms would increase the quality, and quantity of information available to support informed discourse which in turn, would support more informed decision-making. This necessarily involves related knowledge and understandings of what integrating GBA/+ in the evaluation function exist, what enabling environments and tools are required, which barriers are critical to address, and how this information may be used. Key research questions include:

- How do stakeholders identify the relevance and value of integrating GBA/+ in the federal government’s evaluation function?
- What is required from various stakeholders throughout the evaluation process to integrate GBA/+ in the federal government’s evaluation function?
- Where in the evaluation lifecycle can GBA/+ be integrated into the federal government’s evaluation function?

- How does the inclusion of GBA/+ in the federal government evaluation function result in use?

This dissertation presents 6 additional chapters that include the following:

- The theoretical framework for this research and the methodologies employed (Chapter 2 and 3);
- Findings related to the integration of GBA/+ and the federal government's evaluation function, specifically the awareness of requirements, perceived value and benefits, tools and supports, as well as the challenges and mitigation strategies employed (Chapters 4, 5);
- Findings related to the various kinds of evaluation use stemming from GBA+ integration efforts and capacity (Chapter 6)
- Comprehensive findings flowing from assumption-checking against the tailored ToC for integrating GBA+ in the federal government evaluation function as well as various uses of GBA+ in evaluation (Chapter 7), and
- A revisiting of the original research question(s), synthesis of research findings, the implications for possible avenues to future research and professional development (Chapter 8).

Chapter 2 - Dear Betty: The dog ate my recipe card. What ingredients do I need?

This second chapter provides the theoretical background for this research. Research necessarily involves an understanding of intersectionality, and intersectionality theory because of the emphasis accorded by the “+” in GBA+. However, intersectionality alone was deemed insufficient to understanding how a complex phenomenon unfolds within a rapidly evolving policy environment. After defending the research proposal, it was identified that a theoretical framework was required that would embrace complexity. Consequently, the theoretical framework was adjusted to adopt a program theory lens. Program theory, particularly Michie, Stralen & West’s behavioural theory of change, was tailored to address the integration of GBA+ in the federal government evaluation function and is described in greater detail in this thesis chapter.

Intersectionality and intersectional theory

Intersectionality theory refutes single-axis conceptualizations of identity which is important to the evaluation function because of the implications in recognizing that beneficiaries of programs, policies, and interventions are not homogenous. Diverse individuals possess different needs and government interventions’ recognition of these complex needs requires more nuanced recognition, discussion, implementation, and analysis to address these complexities. The identification of these differential needs stems from human rights frameworks in which relationships between categorizations of identity and systems of oppression are prominent. Conceptually, human rights and gender equality are interwoven (UNWomen, 1995; United Nations Evaluation Group, 2014). Gender equality is recognized as essential to human rights, and importantly, serves as a precursor to equality, development, and peace (UNWomen, 1995, p.2). It was in Beijing that gender equality was first acknowledged as a human right (UNWomen,

1995), human rights being defined by the United Nations as “the civil, cultural, economic, political and social rights inherent to all human beings, regardless of one’s nationality, place of residence, sex, sexual orientation, national or ethnic origin, colour, disability, relation, language etc.” (United Nations Evaluation Group, 2011-12, p.11). The *Canadian Charter of Rights and Freedoms* further enshrines the right for individuals to be free from discrimination based on personal identity factors (*Canadian Charter*, 1982, s 6(2)(b)).

The interrelated nature of these personal intersections of diversity and systemic oppression are the critical foundation of feminist intersectionality theory. Distinctly, intersectionality emerged in 1989 from Black feminist legal scholar Kimberly Crenshaw’s work which elocuted that “...women experience oppression in varying configurations and in varying degrees of intensity. Cultural patterns of oppression are not only interrelated, but are bound together and influenced by the intersectional systems of society” (Ritzer, 2007, p. 204). Specifically, Crenshaw drew upon experiences of Black Women in the legal system to demonstrate the limitations of dominant discourses framed by first and second-wave feminisms that emphasized white privilege-based anti-oppression frameworks. She argued that the nexus of gender and race were concomitantly intertwined and unable to be disaggregated from one another. To artificially separate these dimensions results in skewed accounts of discrimination. Additionally, multiple-axis experiences of oppression reinforce and are shaped by one another and the systems that they operate in (Crenshaw, 1989). Feminist intersectionality theory emerged as one response to a revised feminist presentation of gender as the primary axis of identity.

Intersectionality and policy

Despite intersectionality’s emergence as a distinct feminist theory, only recently have the complexities associated with studying intersectionality as a unique approach been identified

(McCall, 2005). Three broad distinctions have been made that categorize intersectional undertakings to include: anticategorical, intercategorical, and intracategorical. (McCall, 2005). These three approaches reside along a continuum. On the left end representing the anticategorical approach, an emphasis is made on dismantling personal categorizations to eliminate discrimination. This point on the intersectionality continuum is fully aligned with transformative approaches to gender mainstreaming. Diametrically opposed on the right end of the scale is the intercategorical approach. Here, acknowledging existing categorizations is integral, and categories are used to “document relationships of inequality amongst social groups and changing configurations of inequality along multiple and conflicting dimensions” (2005, p.1773). Residing in the middle of the continuum is the intracategorical approach. It recognizes the relevancy of categorizations as a means by which to understand complex social interactions. This latter approach to studying intersectionality emphasizes the dynamic nature of these labels and how the individuals who cross label boundaries within the systems that they operate. This point on the intersectionality continuum is aligned more significantly with the integrationist approach to gender mainstreaming discussed in the first thesis chapter. McCall (2005) summarizes that “each approach shares the premise that relationships amongst social groups are containers of definable and indeed measurable inequalities (p.1785). The emphasis in these three intersectionality distinctions shifts amongst structural relationships and relationships of inequality amongst social groups as units of analysis. The taxonomy of these intersectional categorizations wavers between integrationist and transformative approaches to gender mainstreaming.

Recent summaries of intersectional research and analysis efforts identify a historical continuum in which “researchers have traditionally explored intersections of gender and other forms of discrimination as including (1) single axis, (2) comparative, (3) additive, (4)

interactional, and (5) intersectional” (Lewis & Grzanka, 2018, p.2). Several deficits associated with the single axis, comparative and additive approaches exist that are predicated on their lack of attention to the categorizations defined, and explored (Siltanen, 2017). This failure to attend to differential categorizations is important because they are believed to “have importance for identifying configurations of inequality” (Siltanen, 2012, p.4). Given that the diversity of identity dimensions can be expressed at both individual and aggregate levels, research must investigate both. At this juncture, it is unknown as to whether the latter intersectional categorizations and approaches can be truly transformative as gender mainstreaming strategies. It is the definition of anticategorical intersectionality categorizations as defined by McCall (2005) and its refutations of labelling that form the basis of this conclusion.

Despite these analytical forays into intersectionality analysis, there is a dearth of scholarly literature regarding how to systematically undertake intersectional research and analysis. This is methodologically compounded by prominent scholars referencing that there is “no right way to undertake intersectional’ analysis (McCall, 2005; Siltanen, 2012). McCall’s early work on intersectionality distinguishes two forms of intersectionality analysis that includes qualitative accounts of “lived experience of individuals and groups positioned at specific intersections of inequality dimensions, and research on this form of intersectionality is typically qualitative” (McCall, 2005). The other includes predominantly quantitative attempts to intercede in the aggregate, structural level of analysis. “A key focus of an intersectional analysis of this form is to identify which dimensions, and in what combination, are producing the more general pattern of observed inequality” (Siltanen, 2012, p.7). To undertake truly intersectional research and analysis as originally envisioned by GBA/+ advocates within the federal government, advancements away from single axis, comparative simple additive, or comparative approaches

(Hankivsky, Grace, Hunting, Ferlatte, Clark, Fridkin, Giesbrecht, Rudrum & Laviolette, 2012) are required. To undertake truly intersectional research and analysis several principles should be reflected. These principles include intersecting categories; multi-level analysis; power, reflexivity, time and space; diverse knowledges; social justice and equity (pp.28-35).

Intersectionality and evaluation

Hankivsky & Cormier forwarded in 2011 that the development of intersectionality “policy *analysis* (emphasis added) is still undertheorized, and that methods for integrating intersectionality into policy development, implementation and *evaluation* (emphasis added again) are in their early stages of development” (p.217). This holds within the Canadian context: the government’s public-facing GBA+ materials note its alignment with intersectional approaches to policy development, implementation, and analysis, but fail to identify the various ways in which this can be undertaken. While crafting this thesis chapter, guidance materials for the federal government evaluation function on integrating GBA+ in evaluation remained in draft format, and the distribution of these materials was limited. Guidance materials were finalized and released in the summer of 2019.

Despite Hankivsky and Cormier’s (2011) hypothesizing that the analysis dimension of intersectionality being undertheorized, evaluation theorists have been very active in exploring evaluation that is intersectionality-driven. Almost simultaneously to the emergence of feminist intersectionality as a distinctive theory, so too did evaluation theorists offer early knowledge that identified the need to consider not only gender but other intersections of identity in evaluation efforts. Intersectional evaluation theorists echoed feminist intersectional theorist perspectives, citing obligations to address gender, and other elements of diversity (Sielbeck-Bowen, Brisolara, Seigart, Tischler & Whitmore, 2002; Hood & Cassaro, 2002) such as race (Hopson, 1999), class,

and ability (Mertens, 1999; 2014), sexual orientation and religion simultaneously. While many of these early wave intersectional evaluation wave theorists self-identified as feminist (Sielbeck-Bowen, Brisolara, Seigart, Tischler & Whitmore, 2002; Hook & Cassaro, 2002), they were not exclusively so. Amongst these theorists were empowerment (Fetterman, 1994) emancipatory and transformative (Mertens, 1999), and liberatory (House, 1999) advocates with collective social justice ambitions.

Moreover, these evaluation theorists offered multiple applications and mechanisms by which to undertake intersectional evaluation although not referred to as such in the academic literature. These intersectional evaluation theorists support traditionally marginalized groups' engagement in developing and implementing research change agendas to promote social justice. Explicit reasons driving the need for synchronized regard of intersections of diversity were provided. Reverberations of an intersectional feminist theory were reiterated by intersectional evaluation theorists noting "the various combinations of dominant group identities with nondominant group identities create a complex fusion of privilege, power, and disadvantages in each of our lives" (Hood & Cassaro, 2002). It is these authors that provide the strongest example of an anticategorical approach to intersectionality. Paraphrasing Collins on Black Feminism, they reject the possibility of personal categorizations. They note that there is no single viewpoint as the "possibility of multiple trajectories, stemming from common challenges faced by black women, no single, homogeneous, black women's standpoint" (p.8) exist. Other intersectional evaluation theorists on McCall's intersectionality continuum address all continuum sites although an implied acceptance of intercategory, and intracategory is proliferant in the scholarly literature. Fluctuation is evidenced in the intersectional evaluation theorists based upon the epistemological, and ontological practices of the various evaluation approaches. Drawing on

examples from both the legal, and educational systems, intersectionality theorists from both the feminist and evaluation camps, Crenshaw (1989), and House (1999) empirically demonstrated the impact of multiple-axis discrimination using categorizations based on fluctuating sites of race and gender. Ultimately, the communal objectives shared by feminist and evaluation intersectional theorists are not unsurprising given the interdisciplinary nature of, and shared sociological underpinnings of the respective disciplines. Why then, have the significant offerings embedding similar messages from these two distinct various sectors remained largely unexplored and unintegrated?

Mirroring intersectionality theory, intersectional evaluation theorists echoed the need to delve into context. Contextual fluctuations require not only exploring relationships between individuals, groups, and accompanying systems of oppression, but exploring, explaining, and understanding the existing political, cultural, social, economic, and political dimensions of context (Mertens, 1999,2015; Hopson, 1999; Brisolara et al, 2014). These systems have shaped and continue to influence discrimination experienced by individuals. Undertaking analysis to understand identity and identity construction is a particularly complex issue. These systems, like personal categorizations, cannot be dissociated from one another. As intersectionality and personal categorizations are interlocking, so too are the accompanying systems of oppression. Efforts to dismantle one of these systems cannot be undertaken without impacting components of the other systems. Both feminist and evaluation intersectionality theorists note the importance played by these systems in perpetuating discrimination. Mertens (1999) notes three criteria requiring consideration in the application of intersectional analysis. These criteria include an emphasis on the lives and lived experiences of those that have been traditionally marginalized, inequities based on gender, race/ethnicity, and disability are reflected in asymmetrical power

relationships, and social inquiry linked to political and social action (Mertens, Farley, Madison & Singleton, 1994, p.124).

Measuring gender equality and women's empowerment

Intersectionality analysis requires concerted performance measurement and evaluation efforts to situate and identify experienced changes across various margins of personal identity differences within a specific context. This is particularly relevant when considering that the various government departments each engage in multiple and varied evaluation approaches and methods that align with organizational mandates. For example, Economic and Social Development Canada (ESDC) relies heavily on econometric evaluations that parallel their organizational focus in delivering income security-related programs and services to diverse individuals (Economic and Social Development Canada, 2017). In comparison, the Department of Canadian Heritage (PCH) focuses evaluation efforts on reporting on the extent to which inclusion and participation of diverse community members are achieved through third-party delivery of cultural programming initiatives (Department of Canadian Heritage, 2018). Ironically, SWC, the lead agency for gender equality across the federal government, did not apply GBA+ in their evaluation efforts; despite framing intersectionality as an integral theoretical construct in their evaluation of the Women's Program (Status of Women Canada, 2018).

The point of departure for evaluation efforts in their integration of GBA/+ in evaluation necessarily begins with capturing changes on two facets that include a recognition of experienced inequalities, and manifest changes gained as a result of the empowerment experienced by this recognition. For this reason, it is responsible to briefly discuss the evolution of capturing changes in empowerment not only for women but for all those expressing diverse

personal identity elements. The current measurement and evaluation of changes in women's power are grounded in several models stemming from the development landscape in which women are centred as agents of change (Podems, 2010, 2014). Models emerged in the 1970s, and the subsequent decade which are directly responsible for solidifying the relationship in measuring women's empowerment and gender equality. Podems (2010) summarizes these models as including Women in Development (WID), Women and Development (WAD), and Gender and Development (GAD) (p.6). Podems associates evaluation efforts accompanying the development of these models as the inauguration of gendered evaluation. In each of these models, women are situated as the primary unit of analysis, and as the unit of reporting. Derivatives of these 3 models are still employed internationally.

As models evolved to reflect maturing understandings of gender equality, so too has evaluation practice matured to reflect these developments. Kabeer (1999, 2002) notes that the three components critical for inclusion in measuring women's empowerment include resources, agency (both individual and collective), and achievements. It is the introduction of agency at both the individual and collective levels that is critical; without changing the social, political, cultural environment which girls and women inhabit, altering the potential of an individual's agency will ultimately prove unsuccessful. Particularly problematic is the contextual variation. Kabeer (1999) illuminates the role played by choice which is influenced by the confluence of these dimensions and emphasizes empowerment as the changes in the ability to *exercise* choice (Kabeer, 1999).

Akin to intersectionality analysis, symmetrical power relations are inherent to the study of women's empowerment and gender mainstreaming. Specifically, the empowerment of women is how gender equality is achieved. From a feminist scholars' perspective, this extends beyond the

economic domain to address and report on the social, cultural, and political contexts of empowerment as well. Empowerment in these domains are vital dimensions of change and occur at multiple levels including the individual, within relationships and within groups (Kabeer, 1999; Swedish International Development Agency [SIDA], 2002). In this broader sense, empowerment of women specifically involves five components inclusive of: women's sense of self-worth; their right to have and to determine choices; their right to have access to opportunities and resources; their right to have the power to control their own lives, both within and outside the home; and their ability to influence the direction of social change to create a more just social and economic order (UNWomen, 1995, 2015). Empowerment then is both a process and a product (Rowlands, 1997) that is dynamic; and continues to evolve in response to contextual factors. Measuring changes in social norms and mores has been referred to in such ubiquitous terms that include the "art of the impossible" (Langford, 2012); "the distorted metric" (Anderson & Langford, 2012); and "measuring the unmeasurable" (Kloosterman, Benning & Fyles, 2012). Measurement at this level is not an easy undertaking.

Measuring women's empowerment was inaugurated with World Bank economic interests with an accompanying reliance on quantitative indicators (UNWomen, 2015). These quantitative indicators are related to changes in economic statuses (Chant & Sweetman, 2012; Perezneito & Taylor 2014.) This thinking continues to permeate the performance measurement and evaluation landscape despite feminist scholars' offense to framing the measurement and evaluation of gender equality solely through women's economic empowerment (Sweetman & Chant, 2012; Perezniето & Taylor, 2014) for several reasons. Objectifying women for their potential capacity to contribute to the labour market is problematic as the value of women's contribution is limited to their ability to contribute financially. Limiting women's empowerment to the economic

spheres also fails to address barriers to gender equality in social, political, and cultural dominions. This tension fails to address caregiving responsibilities traditionally borne by women that include familial and community obligations. These scholars suggest that the social and political terrains influencing labour market conditions are omitted from change agendas, thus limiting the achievement of gender equality for its intrinsic value. Ultimately, feminists argue that advancing gender equality for its intrinsic value is sufficient.

This approach promoted by the World Bank frames investing in women as *smart economics* (WorldBank, 1994). “Investing in women is critical for poverty reduction. It speeds economic development by raising productivity and promoting the more efficient use of resources. It produces significant social returns, improving child survival and reducing fertility. And it has considerable intergenerational payoffs” (p.17). Women’s capital is seen as a resource that can be profited from. Moser (1989) coined this “the efficiency approach” to describe women’s economic prioritization in development programming and resulting measurement efforts. Many development agencies, including the United Nations (UN) family of organizations, perpetuate rationales driving smart economics. These agencies have demonstrated through performance measurement and evaluation efforts that the return on investment (ROI) made from dedicating resources to addressing barriers faced by women, extends proportionately to their children and communities. Women who are economically empowered as a result of their increased access to, and ability to participate as agents in the labour market then extend these benefits to their children and the broader community. This phenomenon has been christened “the double dividend” (United Nations Children’s Fund, 2007).

Gender-based Analysis Plus (GBA+), intersectionality and evaluation

Status of Women Canada's GBA/+ materials note that "the plus in GBA+ acknowledges that GBA goes beyond biological (sex) and socio-cultural (gender) differences. We all have multiple identity factors that intersect to make us who we are; GBA+ also considers many other identity factors, like race, ethnicity, religion, age, and mental or physical disability" (SWC, 2017). The "plus" in GBA+ was added in 2010-11 fiscal year during an organizational ministerial transition; updating previous guidance materials (Status of Women Canada, 1998) to include multiple personal identity elements. Further, SWC's online GBA+ training refers to an intersectional analytical process "to ensure that *all* aspects of diversity are considered when an initiative is analyzed" (Status of Women Canada, 2018) although training materials have been updated to highlight *some* elements of diversity. This directly conflicts with advice provided on intersectional analysis (Hankivsky, 2011; Siltanen, 2012).

GBA/+ training materials have sections dedicated to various functional communities within the federal government including research, policy development, program design, legal counsel, and communication. In these training materials, evaluation and monitoring (or performance measurement) functions have been grouped into one functional community. This is problematic as lines of responsibility are generally allocated amongst different Assistant Deputy Ministers (ADMs) in federal organizations. These blurred accountability requirements require close coordination which is difficult when each function is operating at optimal levels. Numerous Treasury Board Secretariat studies of the evaluation function (2004a, 2004b) bemoan the deficit of sufficient performance measurement (monitoring) information in undertaking evaluations. This is echoed in academic studies of the federal government's evaluation function including specific studies of evaluation recommendations and strategic evaluation use (Bourgeois & Whynot, 2018a, Bourgeois & Whynot, 2018b, Segsworth, 2005). GBA/+ training materials task

the performance measurement and evaluation functions to consider the following questions (SWC, 2018):

- Were diverse perspectives assessed and factored into the initial design?
- Does the sex- and/or gender-disaggregated data being used in the evaluation phase include diverse people?
- Have baseline indicators been established to measure the effectiveness of the initiative, and are they conducive to assessing the impact on various sub-groups of the target population? If not, what data collection methods would best measure outcomes for these groups?
- Are there gaps in the quantitative or qualitative data needed to effectively measure outcomes? How could these gaps be filled?
- Are there differences in how various groups access or experience the program or service? Are there gaps that pose risks or create unintended barriers for some groups? How can these gaps be filled?

While these questions are indeed relevant for the federal government's performance measurement and evaluation functional communities, the majority of questions posed reside directly under the responsibility of program management. Many of the previously introduced policy cycle junctures now substantively requiring GBA+ information (such as Budgets, MCs, and TB Subs) to occur before interventions begin programming activities. Federal government evaluators may or may not be involved at this design juncture; this does not preclude evaluators from participating in program design and delivery discussions despite the focus on ex-post perspectives. Of the questions posed in the above list, in an ex-post application, only the latter two of the questions posed would align themselves directly. While the other questions posed

could conceivably be included in an evaluation as either sub-questions or, reframed into indicators, the inclusion of other questions would be a waste of scarce evaluation resources. This is discussed in greater detail under the thesis chapter dealing with challenges and mitigation strategies.

Changing theoretical framework direction

After defending the research proposal and entering the data collection research phase, it was identified that intersectionality as a conceptual framework alone was insufficient to support this research. While intersectionality and GBA/+ identified *what* was to occur, the requirement for more detailed elaborations of *how* this was to occur was identified. Options considered included the Bennett Hierarchy based on knowledge, aspirations, skills, and attitudes, acquisition and change (KASA), Sen's (2004) and Nussbaum's (2011) Capability Approach (CA) with strong ties to policy and structural changes, and, Michie, Stralen and West's (2011) COM-B theory of behavioural change, and Mayne's adaptation of the COM-B behavioural theory for program theory were among the forerunners. The CA was ultimately not selected for two reasons that included its normative orientation; "it is not a theory that will *explain* poverty, inequality, or well-being, but rather a theory that helps us to conceptualize these notions" (Robeyns, 2016). Bennett's Hierarchy was similarly rejected for its application failure in explaining how changes might occur.

Program theory and theories of change (ToC)

Theories of change (ToC) are "essentially a comprehensive description and illustration of how and why the desired change is expected to happen within a particular context." A ToC is outcome-focused; highlights causal attribution (Clark & Anderson, 2004); and assumptive insights that accompany movement from one phase in the results chain to another. Scriven

defines a theory of change as a hypothesis about the way that a program brings about its effects (1991). It is the *logic*, not to be confused with a logic model, impact pathway, causal attribution results chain, or any of the other synonymous terms that are often used interchangeably, that an initiative uses to postulate about how potential changes will occur as a result of undertaking targeted activities. Federal government evaluators are very familiar with ToC; 5 out of 6 federal government guidance documents for the evaluation function specifically situate theories of change as integral to understanding interventions (Whynot, Lemire & Montague, 2019).

ToC are understood to be succession-based; the arrows within the results chain are employed to note either a progression or regression along the results continuum. When a preceding results phase and associated conditions are met, then advancement will occur. If preceding outcomes and assumptions are not met; double-ended arrows note the potential for “backslide”. All results phase outcomes and assumptions usually outlining contextual conditions, have to be met before the ultimate objective or goal of the intervention is achieved. The results chain continuum in its entirety explains what will be achieved, and with the accompanying articulation of associated assumptions, unpacks what happens between what has been referred to as the “black boxes” of the results chain. In this way, ToC are focused in particular on “mapping out” or “filling in” what has been described as the “missing middle” between what a “...change initiative does and how these leads to desired goals being achieved” (Centre for Theory of Change, 2019).

Embedded within ToC are explanatory mechanisms. These explanatory mechanisms used can be traced back to the formal work of Donald Kirkpatrick (1959), Edward Suchman (1971), and Daniel Stufflebeam (1967), among others. Program theory as a distinct theoretical framework flourished with the formalization of theory-based evaluation (TBE) in the 1980s

(Chen and Rossi, 1980, 1983; Bickman, 1987). In the last 2 decades, a wealth of resources dedicated to ToC and program theory have emerged. These include reflections on different types of program theories (Shadish, 1987), what constitutes good program theory (Davies, 2012; Mayne, 2015), how and with what purpose to use theory (Stewart, 1997), how to develop theory (Funnell and Rogers, 2011), and how to test and use program theories (Bickman, 1987; Mark, 1990; Rogers et al., 2000).

The benefits of ToC are numerous and varied; they can be used pragmatically, strategically, and operationally at multiple levels and with various stakeholders. They can be applied across all intervention stages including design, for implementation and monitoring, at evaluation, and at intervention closure (Dhillon & Vaca, 2018). They are particularly useful in differentiating between assessments of success and failure and “uncovering what aspects of an intervention have or have not been successful with which particular subgroups” (Blamey & MacKenzie, 2005, p.151). The introduction of subgroups was particularly important for this research given the need to integrate expertise from three of the federal government’s functional communities including evaluation, performance measurement, and gendered expertise. Three quality criteria for ToC were originally identified that included plausibility, do-ability, and being testable (Connell & Kubisch, 1998) although associated quality criteria employed in assessing ToC (Davies, 2012; Mayne, 2015) have evolved to mirror ToC themselves. Additional quality criteria identified as important by Davies and Mayne include understandability, verifiability, explainability, completeness, inclusiveness, justifiableness, owned, embedded, weighted, and purpose-driven.

Social science theory should be a mainstay of ToC (Whynot, Lemire & Montague, 2019). Centrally locating social science theories within ToC leads users to have confidence in the

multiple ways in which ToC are used, and the validity/reliability of these uses is increased. Both social science and program theories have been employed predominantly by evaluators, and program managers to assist in identifying what successes have been achieved and learning from implementation failures. However, program theories have an intuitive appeal to a much broader audience that should include performance measurement, policy, and research functional communities (Whynot, Lemire & Montague, 2019). Recently, Brousselle and Buregeya (2018) have set the stage for a rise of a new generation of TBE reaching to address the challenges and limitations presented within current program theory thinking to address complex social issues, and the associated complexity grounded in open systems which are in turn, embedded in multiple social systems (Brousselle & Buregeya, 2018, p.163-164). This necessarily involves the engagement of multiple actors who undertake various roles and responsibilities in delivering complex interventions (Koleros, A., Mulkerne, S., Oldenbeuving, M. & Stein, D.; 2018; Koleros & Mayne, 2019). Integrating GBA/+ in the federal government's evaluation function is imbued with complexity involving a vast array of diverse stakeholders, a shifting policy landscape with accompanying demands, and changing legislative requirements. Consequently, for this research, a theoretical framework was required that was sufficiently flexible enough to deal with these conditions but was also able to adeptly respond to key complexity questions dealing with the uncertainty created by interdependency among numerous actors that are creating and adapting. That program theory can do so ensures its alignment as a theoretical framework for this research.

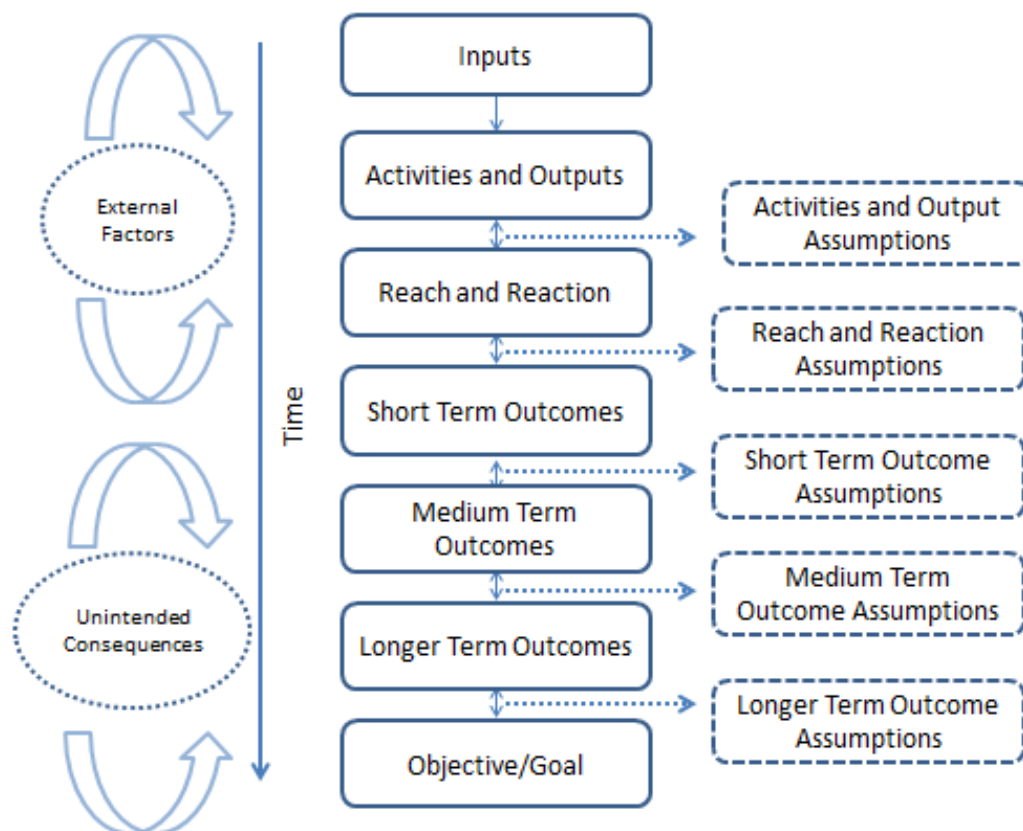
Early ToC were very similar to logic models; critiques focused on simplistic depictions and the failure to address what was occurring during implementation. Early main categorizations of ToC dimensions were simplistically underscored to include:

- *inputs* (which are the resources required to implement an intervention);

- *activities and outputs* (what the intervention will do and what is produced);
- various *levels of outcomes* including short, medium, and longer term (the results that occur as a result of undertaking activities and producing outcomes); and
- the intervention's *impact* (why an intervention is being undertaken).

As ToC evolved, additional dimensions were flagged as necessary for inclusion. This included reach and reaction (Montague & Porteous, 2013) which are defined as “the target that a given program or organization is intended to influence including individuals and organizations, clients, partners, and other stakeholders” (Montague & Porteous, 2011, p.6). Similarly, Mayne (2016, p.5) defines reach/reaction as “target groups who are intended to receive the intervention's outputs and their initial reaction” (Mayne, 2015, p.5). The incorporation of reach and reaction as a dimension of the tailored ToC designed for this research was particularly appealing given the varying epistemological backgrounds of the 2 mechanisms, and associated stakeholders identified to participate in this research. Additional dimensions deemed essential included: external factors including enabling and hindering influences; unintended consequences, and timelines (Mayne, 2015). Practitioners using ToC have embraced the addition of these dimensions, ensuring that intervention complexities are adequately addressed. Figure 2 presents a generic ToC used as the foundation for this research.

Figure 3 Generic theory of change and associated dimensions



The behavioural change theory of change (ToC)

Policy is inherently messy. “Wicked policy problems are characterized by conflicting values and perspectives, uncertainties about complex causal relationships, and debate about the impacts of policy options” (Head, 2018). This broad characterization of policy certainly applies to Canada’s policy cycle, and the position of the federal government evaluation function compounds its interpretation of wickedness. Federal government policies, programs, processes, and interventions are influenced and operate under the jurisdiction of a wide spectrum of various schools of thought influenced by various epistemological backgrounds. Darnton’s (2008) review of behavioural change for policy identified multiple backgrounds and factors including economics; behavioural economics; information and value action gaps; values, beliefs, and

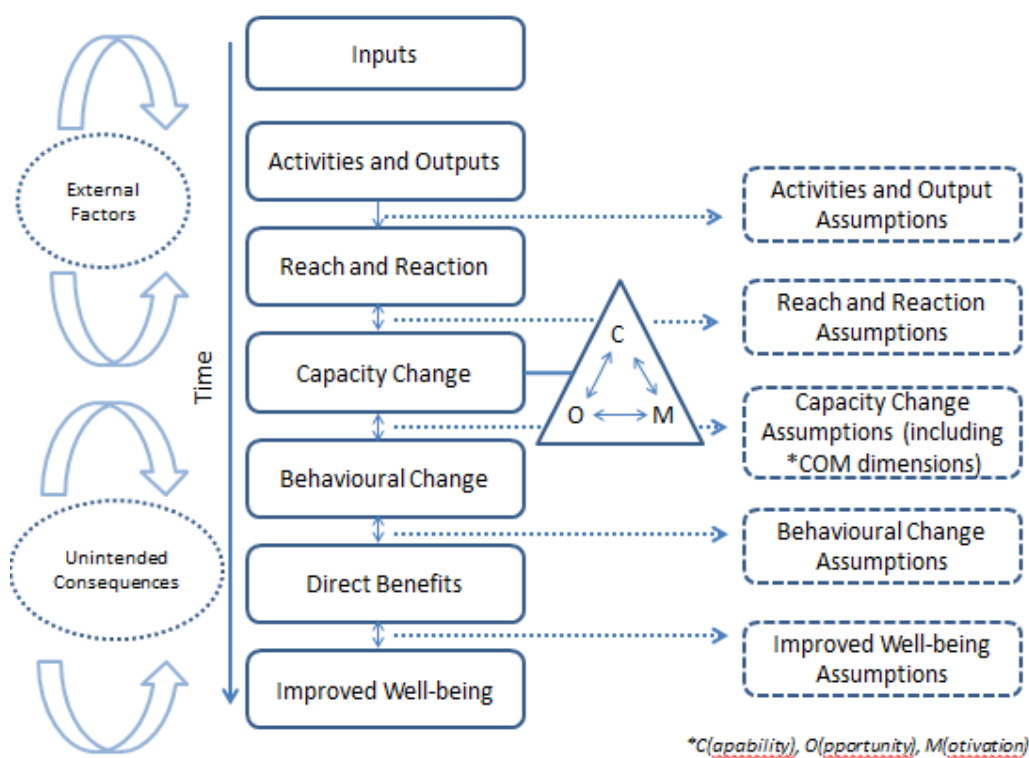
attitudes; norms and identity; agency, efficacy and control; habit and routine, the role of emotions; external factors; self-regulation, and societal factors (Darnton, 2008). Desired behavioural change directed and influence by various policies occur at multiple levels, involve multiple actors and operate within and between systems. How then does one delve into the intricacies of how and why behavioural changes occur? Darnton's review concluded that:

The review of behavioural models and theories reveals human behaviour as complex, arising from diverse psychological factors, and from social, societal and contextual influences. Building on standard economic assumptions, social-psychological theory shows people's motivations to be wider than self-interest, and to be varying over time, and in different contexts. Most importantly, theory reveals behaviour to be both more and less rational, sometimes strongly influenced by emotions, or habits and routines. The diversity of factors at play in social-psychological models explains why changing behaviours has proven so challenging for policy makers. Interventions must address a number of factors at once, and be flexible to different audiences and contexts (2008, p.68)

Embedding evaluation in this framing of behavioural change within the policy landscape becomes ultimately even more complex. Contributions from behavioural change theorists in developing and implementing corresponding ToC for complex interventions cannot be ignored. Michie, VanStralen & West (2011) adopted this complexity lens and forwarded an intuitive behavioural change theory that structures behavioural change occurring based on the mutually reinforcing interaction between three conditions that include Capabilities (C), Opportunities (O), and Motivations (M). This model identifies that behavioural change (B) occurs *only* when these three dimensions are present. This is the COM-B behavioural change model. Should any one of these three dimensions not be present, behavioural change will ultimately not occur. These definitions are very similar to dimensions identified in the Bennett Hierarchy including Knowledge (K), Aspirations (A), Skills (S), and Attitudes (A). It is the insertion of the Opportunity (O) dimension that differentiates between these two models of behavioural change.

Mayne (2015) builds on West, Michie and VanStalens's behavioural change model, integrating behavioural change dimensions into theories of change. This is the COM-B theory of change. Highlighted below in Figure 4.

Figure 4 Generic COM-B theory of change



Understanding the COM-B theory of change resides in the pivotal C(apacity) change dimension, which as previously mentioned, is cumulatively comprised of C(ability), O(opportunity), and M(otivation). Capacity is the mid-point of the ToC. This main difference between Michie, VanStralen & West and Mayne's view of the COM-B ToC resides with Mayne's embellishment of M(otivation) "it includes habitual processes, emotional responding, as well as analytical decision-making" (Mayne, 2016, p.4). Mayne's focus on analytical decision-making is aligned with the objectives of both GBA/+ and the evaluation function in their provision of information to decision-makers.

Few accounts of the specific explorations between gender mainstreaming and ToC are available in this burgeoning field of scholarship. Only recently have practitioner-based efforts regarding gender and program theory/ToC begun to coalesce amongst development practitioners such as those authored by VanEerdewijk and Brouwers (2015) on behalf of HIVOS, Jost, Kristjanson and Ferdous's contribution (2015) to Climate Change, Agriculture and Food Security, and the Health Communication Capacity Collaborative (2018). The emergence of related policy briefs and guidance materials emphasize the importance of including gender in program theory/ToC regardless of the interventions' intended impact. Newly minted policy briefs and guidance on integrating gender in program theory (VanEerdewijk and Brouwers, 2015) reference program theory as a mechanism for facilitating understanding and creating opportunities to facilitate understanding from various situated knowledges. These materials are helpful in their identification of social and behavioural determinants impacting how individuals engage in various levels of the systems in which they are embedded (Health Communication Capacity Collaborative, 2018). With the intentional emphasis on gender transformative results embedded in the Sustainable Development Goals (SDGs) agenda, using gender in program theory/ToC represents an opportunity to circumvent some of the challenges historically encountered in gender mainstreaming efforts (Whynot, Mavriplis, Farenhorst, Langelier, Franz-Odendaal, & Shannon, L, 2018).

Development of a tailored theory of change for this research

At its broadest, the COM-B ToC is one used to explore behavioural change at individual, organizational, and systems levels. The federal government evaluation function represents a system comprised of both individuals, and organizations responsible for evaluation policy and practice implementation. Together these components constitute the federal government

evaluation function. Ostensibly, each of these organizations contributes to the federal government evaluation system but also operates within their organizations as unique systems. On occasion, several organizations may work together at a sector level. The federal evaluation system has a complex web of actors involving representatives from program areas, policy, planning, performance measurement, gendered expertise in addition to evaluators as well as subject matter experts. These actors influence not only the federal evaluation policy but in turn, also influence organization evaluation and sector policies. The purpose of this research is not to map system actors, but rather to recognize the necessary representation of a diverse array of constituents from various functional communities contributing to the successful integration of GBA+ in the federal government evaluation function.

Following a shift in theoretical frameworks, a specific ToC regarding the integrating GBA+ in the federal government evaluation function was drafted to support this research. It draws on the best available evidence drawn from national and international sources, as well as academic content. The tailored ToC references accounts of the implementation and perceived failure of GBA/+ specific to the Canadian context. Many of the findings drawn from these suggested best practices are reflected in the identification of associated assumptions specific to particular results phases presented in the tailored ToC. Assumptions include both general (those related to gender mainstreaming/GBA+ theory) and specific (those related to organizational details including size and service delivery focus). Early draft ToC change content was shared with program theory developers and influencers for feedback including John Mayne (personal communication, 2017) who posed several questions for consideration that improved the original draft clarity including questioning what role was played by SWC, particularly concerning the

evaluation function. Presented below in Table 1 includes content from the tailored ToC developed.

Table 1 Tailored theory of change (ToC) for integrating Gender-based Analysis Plus (GBA+) in the federal government evaluation function

COM-B Theory Component	Theory of Change General Assumptions	Organizational Evaluation Assumptions
Activities and Outputs - GBA+ policy cycle requirements are communicated - Junctures at which evaluation can contribute are identified	A1. SWC (lead agency) for GBA+ has capacity/expertise A2. Federal government priority of GBA+ recognized by organizations	B1. Evaluation functions aware of GBA+ policy requirements B2. Evaluation functions associated with International Affairs (IA) and Social Affairs (SA) sectors more aware of GBA+
Reach and Reactions - Evaluation function recognizes GBA+ requirements - Organizations are positively engaged in applying GBA+ to evaluation - Functions contributing to evaluation are supportive/supported in integrating GBA+	A3. Organizations realize communications are directed at them A4. Organizations recognize relationship between gender equality and GBA+ A5. Organizations recognize value of incorporating GBA+ in evaluation A6. Central agencies provide related leadership for other functional areas	B3. Evaluation functions aware of <i>Policy on Results</i> requirements B4. Functions supporting evaluation would be aware of <i>Policy on Results</i> requirements
Capacity Change Stakeholders knowledgeable about applying GBA+ in evaluation	A7. Organizational tools/supports are developed that support integrating GBA+ in evaluation A8. Organizational functions coordinate to support integrating GBA+ in evaluation	B5. Evaluation functions aware of SWC's GBA+ online training B6. Evaluation functions engaged with Gender Champions (GCs) and/or Gender Focal Points (GFPs)
Motivations - New Policy on Results provides framework for integrating GBA+ in evaluation - Compliance with central agencies requirements to integrate GBA+ at other junctures in the policy cycle	A9. Continued political leadership A10. Department/agency leadership prioritizes gender equality	B7. Evaluation functions influenced by policy B8. Evaluation functions comply with policy requirements B9. Evaluation functions in SA and Large Departments or Agencies (LDAs) would be comfortable with GBA+ in evaluation B10. Small Departments or Agencies (SDAs) would not be comfortable applying GBA+ in evaluation
Opportunities Commitment to gender equality and GBA+ is recognized by organization and function leadership	A11. Continued government commitment A12. Continued political leadership A13. Organizations systematically embed GBA+ in evaluation and other government functions	B11. Evaluation functions supported by organizational and functional leadership B12. Evaluation functions do not have necessary financial resources in place

Capabilities Organizations access and/or create training opportunities to build individual and institution practice changes	A14. Training opportunities available A15. Knowledge can/is applied within individual organizations A16. Other functional areas provide coordinated support/efforts to evaluation in individual organizations	B13. Evaluation functions take SWC's GBA+ online training B14. Evaluation functions access guidance materials B15. Evaluation functions take and/or create GBA+ training specific to results context B16. Evaluation functions with GBA+ training have higher capabilities B17. Programs with embedded elements of diversity apply GBA+ in evaluation
Behavioural Change Federal government departments/agencies apply GBA+ in their evaluations	A17. Sustainable political and government leadership in GBA+ A18. Organizing functions coordinate to support integrating GBA+ in other junctures of policy cycle A19. Integrated/comprehensiveness of GBA+ across policy cycle A20. Opportunity costs related to GBA+ don't exceed expectations	B18. Evaluation functions do not have access to disaggregated data by identity dimension
Direct Benefits Decision-makers use GBA+ information to inform policies, programs, and services	A21. GBA + information is presented in evaluation reports A22. GBA+ information is considered in decision-making A23. Value of GBA+ is apparent	B19. Decision-makers consider and request GBA+ in evaluation frameworks B20. Decision-makers use GBA+ information stemming from evaluation processes and products
Improved Well-being Improved equity, diversity, and inclusion		

Tailored ToC content presented above was finalized using various lines of evidence associated with this research. A final product and accompanying visual representation and narrative are consequently presented in the final thesis chapter of this dissertation. ToC theorists (Dhillon & Vaca, 2018) have suggested that theories of change “often include far less information that they potentially could and that the inclusion of these additional elements could add a substantial benefit” (p.78). It is strongly suggested that ToC have a narrative accompany the graphic depiction (Mayne, 2015). What follows in this paragraph is a specific narrative accompanying the ToC. The federal government indicates that gender equality, and GBA/+ as a mechanism supporting gender mainstreaming, is a priority. As such, federal government organizations recognize their associated responsibilities to implement GBA/+ in their work, including within the evaluation function. This necessitates all sectors across government to recognize the relevance of GBA/+ within their domains. Evaluators can play a key role in challenging assumptions that perpetuate discrimination based on personal identity factors and have the necessary capacity to do so. The development of this (C)apacity applies to both individuals working in proximity to the evaluation function, and within the evaluation function at conceptual and practical levels. The (O)ppportunity dimension is represented by the political leadership of Justin Trudeau who has self-identified as “feminist” and has been an advocate of gender equality and GBA/+. It is anticipated that positioning Justin Trudeau’s leadership as an (O)ppportunity within this conceptual framework will assist in explaining the advancement of gender-based analysis but will also highlight sustainability requirements. The (M)otivation dimension is represented by the Policy on Results’ (2016) articulation of GBA/+ as a priority, as well as other policy junctures requiring gender-based analysis information such as Budgets, TB Subs, and MCs. Cumulatively, the dimensions that comprise the (C)apacity dimension will result

in evaluations applying GBA/+ in their evaluation processes, and products. Importantly, this will involve the coordination amongst a diverse array of stakeholders to support these efforts.

Information resulting from the integration of GBA/+ in evaluation efforts will be used to inform decision-making and implement changes to improve the well-being of all Canadians.

Simplistically this narrative reflects a systems approach to a whole-of-government shift in how GBA+ is integrated in evaluation efforts from awareness through to utilization. Findings from the various lines of evidence associated with identified results phases and assumptions identified in this ToC are discussed in the following thesis chapters.

Chapter 3 – Dear Betty: I know that proportions are important. How much of each ingredient?

This thesis chapter recognizes that sexism occurs in the research process and discusses efforts employed to mitigate the impact of this phenomena across design, conduct, and analysis efforts. A contribution regarding my positionality relative to the federal government’s evaluation function and expertise in gender mainstreaming is included. The chapter highlights the benefits of mixed-methods research and details the qualitative and quantitative lines of evidence employed in exploring the integration of GBA+ in the federal government’s evaluation. These lines of evidence included:

- A critical review of the literature identifying the linkage between GBA+ and the federal government evaluation function;
- Case studies conducted with federal government evaluation functions integrating GBA+;
- A survey of those individuals responsible for the evaluation function (Heads of Evaluation [HoE]); and
- Key informant interviews with representatives from the government, evaluation practitioners including feminist evaluators, and academics.

The chapter concludes with a brief examination of how analysis and reporting efforts associated with this research were approached.

Sexism in research

Early feminist researchers such as Eichler (1987) identified four dimensions potentially producing sexism in the research process that included: “access by participants to research processes; the conditions under which research is conducted, the research process itself, and how research results are reported and published” (p.3). It is with these junctures in mind that I

designed this research to avoid sexist language, sexist concepts, the adoption of androcentric perspectives, sexist methodology, and sexist interpretation of results to the extent possible.

Further, I acknowledge that my research occurs within systems and institutions that are sites of the perpetuation of discrimination in the power that they hold and the controls that they exert. In this manner and paraphrasing Audre Lorde's famous disassembly quote, any subsequent changes will result only in the "most narrow perimeters of change possible and allowable" (Lorde, 1979). Structurally, the research approach identified will likely not result in transformative changes in equity but rather at best, only small, incremental steps if any, towards change. My research builds on historical contributions and lays the foundation for subsequent scholars both theoretically, and in implementation.

Positionality

Within this research, it is integral that I position myself. As a returnee to academia after almost twenty years of providing performance expertise for both the federal government and within the non-profit community, I occupy a leadership position within this functional community which is discussed below. I have been employed by two of the key government departments bearing responsibilities relevant to this research that include Treasury Board Secretariat (TBS) who is responsible for the federal government's evaluation policy and its implementation, as well as SWC who is the agency responsible, and the federal government's lead for gender equality and the implementation of GBA/+ across government. Where these two organizations historically intersected, and where I became interested in my dissertation topic, was in the production of guidance materials on integrating GBA/+ in evaluation following the release of the first OAG report in 2009. I had served on an oversight committee on the first concerted effort to develop materials in 2010 during the transition from GBA to GBA+, and later

in 2015 as the first author on the guidance materials before the release of the 2016 *Policy on Results*. Additionally, I had served as the temporary HoE within SWC before leaving the organization to attend full-time doctoral studies. During the case study data collection period, it should also be noted that I was employed on a casual basis as the Principal Analyst at the Centre of Expertise for GBA+ at Economic and Social Development Canada (ESDC), and as a senior evaluator at the Natural Sciences and Engineering Council of Canada (NSERC).

During my tenure within the federal government evaluation function community, I have developed both professional and personal relationships with a significant portion of senior individuals including Heads of Agency (HoA), Assistant Deputy Ministers (ADMs), Director Generals (DGs), Directors, and managers responsible for both gender equality and evaluation. I have served federal government evaluators in the National Capital Region (NCR) as one Past-President of the National Capital Chapter of the Canadian Evaluation Society (NCC-CES). I am additionally responsible for nurturing a portion of emerging federal public service evaluators in my role as a contract instructor in Carleton University's graduate certificate in program evaluation. Internationally, I serve on EvalGender+ (a group that supports the inclusion of gender and other identity factors in evaluation) and have provided input into the Sustainable Development Goals (SDG) agenda renewal including how gender equality should manifest and across the goals, as well as having provided input into their measurement. I also served as the Program Co-Chair for the American Evaluation Association's (AEA) Topical Interest Group (TIG) on Feminist and Gender Issues in Evaluation.

This rich combination of experiences provides unique insight and opportunities, and also serves to disadvantage me as I am a product of, and influenced by my experiences. I have been a long-standing advocate for the inclusion of gender in evaluation processes and products. In this

research, I am both an insider who is knowledgeable about the issue, members of the community, and how it operates as well as an outsider researching a functional community that has been, and will continue to be a professional area of practice once academic studies are completed. The implications of the adoption of both insider and outsider roles manifest in the negotiating of boundaries between academic conceptualizations and practitioner implementation of GBA/+ within the evaluation function.

The case for mixed methods

The federal government evaluation function is embedded in positivist foundations whereas GBA/+ orients itself firmly in a post-positivist research paradigm in its attendance to social constructions and multiple axial expressions of identity. In the positivist paradigm, a reliance on quantitative methods and deductive reasoning support the production of evidence; whereas, in the post-positivist schools, qualitative accounts are acceptable and inductive approaches common. Early efforts to systematically embed GBA with its feminist underpinnings amidst the bureaucracy's construction of "neutral" evidence resulted in negative reactions, and even blatant disregard of official requirements (Office of the Auditor General, 2009, Office of the Auditor General, 2015). GBA/+'s potential as an outcome reporting mechanism has been shortchanged in its effective deployment across governmental performance stories (Bustelo, 2016) often to the detriment of those with multiple identity factors and considered to be marginalized. As feminist methodologist Eichler notes (1987), these research paradigms become entrenched over time and are resistant to change.

The purpose of my research is not to identify one paradigmatic preference over another, but rather to emphasize its practical application in what works well with the stakeholders involved. Consequently, a mixed-methods approach as identified by Greene (2007) was

employed for this research, drawing on both qualitative and quantitative methods to “legitimize multiple rationalities for and multiple ways of practicing social inquiry” (p.42). It employs both deductive and inductive approaches. The stakeholders involved represent various perspectives from both research paradigms. For example, gender expertise in the form of GFP and GC, as well as EDI specialists, could be associated with post-positivists schools of thought.

Alternatively, evaluation and/or performance measurement representatives could identify with positivist constructions. Important chasms currently exist amongst key stakeholders involved in this area, particularly between the different government function areas. Neither gender specialists, nor evaluators bear responsibility for ensuring that gender equality, and GBA/+ is considered in the reporting of program outcomes. While gender equality is a shared responsibility, the performance measurement and evaluation dimension related to gender equality reside with program management. None of these government functions (gender specialists, evaluators, nor program management) are directly responsible for ensuring that sufficient disaggregated data is available to support evaluation, this effort is jointly undertaken and positions the program as the lead with performance measurement functions assuming a supporting role.

Consequently, the most significant praxis consideration is the necessity of coalescing these diverse groups’ efforts under a common agenda of improving the quality of the information provided to Canadians, Parliamentarians, and senior government officials to improve the development, renewal, and assessment of programs. In doing so, it is believed that ultimately the lives of Canadians will be improved as a result of reduced inequality. My research explores the necessary collaboration amongst various stakeholders to identify which stakeholders need to be

involved, where they need to be involved, and actions required to improve both effectiveness and efficiencies of processes and resulting products when integrating GBA/+ in evaluation.

Within an evaluative context, UN Women notes a historical reliance on qualitative methods to report on women's empowerment, and gender equality (UN Women Asia-Pacific, 2014). This inclination unfavourably impacts perceptions of an evaluation's rigour and validity. Similarly, UNICEF notes that evaluation findings not supported by both qualitative, and quantitative lines of inquiry are compromised by perceptions of weak rigour, or compromised validity (2011). In many ways, the adoption of a mixed-methods approach is well aligned with GBA/+ and intersectional research itself, drawing on, and requiring both qualitative and quantitative information. Mixed-methods methodologies refer to "an appropriate mix of qualitative and quantitative methods ... to gather and analyze data in order to offer diverse perspectives ... and to promote participation of different groups of stakeholders" (United Nations Evaluation Group, 2011-12, p.38). Several mixed-methods model designs are highlighted that include sequential, parallel, and multi-level options (United Nations Children's Fund, 2011, p.54). Support for mixed-methods is unambiguous. Several UN organizations, including UNICEF, extensively outlines the associated benefits of supporting the use of both qualitative and quantitative methods throughout research (2011-12). The benefits of employing a mixed-methods methodology is substantiated in numerous sectors; particularly in gender equality (Pereznieto & Taylor, 2014). Employing a mixed-methods approach extends the possibility of "mak[ing] statistically representative generalizations to the wider populations" (United Nations Children's Fund, 2011, p.51) which will be helpful to other public administration settings considering implementing similar initiatives within their evaluation functions.

My research was initiated with a deductive approach framed by the tailored COM-B theory of change regarding the integration of GBA+ in the federal government evaluation function. This tailored COM-B ToC was developed using identified dimensions drawn from international best practices on gender mainstreaming, as well as academic literature reporting on the state of GBA+ implementation in Canada. First drafts of the tailored ToC were tabled for discussion and feedback with evaluation theorists as well. The tailored ToC itself was tabled for feedback from evaluation peers including the Planning and Performance Exchange Annual Symposium (Whynot, 2018), the European Evaluation Society's Bi-annual Conference in Thessaloniki Greece (Whynot b, 2018) as well as the American Evaluation Association's Annual Conference in Cleveland, Ohio (Whynot c, 2018). Audience suggestions from these venues served to clarify results phases, fine-tune assumptions in terms of organizational size and sector, and highlight additional assumptions that were not originally included in the original draft of the tailored ToC.

My research adopted a multi-level mixed methods research design that embeds the tailored ToC across all related data collection, analysis, and reporting dimensions. Each data collection tool included elements from the tailored ToC. Careful consideration of where each stakeholder would potentially be able to contribute relevant data at specific junctures was provided. The research was instigated with a critical review and was followed by the initiation of case studies as a line of evidence examining how federal government organizations are approaching this integration. Key informant interviews were interspersed amongst case study efforts and was followed by a survey of the HoE. Following the HoE survey, data collection efforts concluded with additional key informant interviews. As data became available from the various lines of evidence, I refreshed the tailored ToC to ensure its currency in light of the dynamic policy

landscape. Findings from each of these lines of evidence are discussed within subsequent dissertation chapters.

The tailored COM-B ToC content was finalized with findings from various lines of evidence associated with my research. The final draft is presented in graphic format, in conjunction with an accompanying narrative in the concluding chapters of this dissertation. My purpose is to represent visually for readers the complexity of what constitutes the integration of GBA+ in evaluation, and in addition, provide a background for future academics and practitioners who wish to build on studying GBA+ efforts using a ToC approach.

Lines of evidence

Critical review

One form of systematic review underutilized as a line of evidence in my research is the critical review (Grant & Booth, 2009). This type of systematic review differs from traditional understandings of what has been differentially presented as literature, systematic, meta-analysis, umbrella, and rapid review forms. What differentiates this form of systematic review is its extensive explorations into a specific topic by using significant resources in the field. This involves the development of an accompanying conceptual or chronological narrative. The benefits of engaging with available resources in this manner focus on its potential theoretical contribution to existing, or new ways of thinking. Findings from these efforts can be used to inform both qualitative and quantitative inquiry (Bretschneider, Cirilli, Jones, Lynch & Wilson, 2017). I selected this line of evidence because of the distinct timelines associated with the integration of GBA/+ in the Canadian federal government evaluation function, and the limited number of related resources available.

I harvested critical review resources from a variety of sources including publicly available websites, materials provided as part of GBA/+ training that are several decades old, and from the Public Works and Consulting Services Canada (PWCSC) library branch which serves as the repository for all government historical documents such as policy circulars and other policy documents. Authorship of resources featured in this critical review was isolated to various government organizations including the Standing Committee on the Status of Women (FEWO), Status of Women (SWC), Office of the Auditor General (OAG), Treasury Board Secretariat (TBS), as well as the Royal Commission on the Status of Women (Byrd Report), and available materials on GBA or the integration of GBA in evaluation materials by individual departments. The total number of documents included in this critical review by author and year are presented in Table 2 below.

Table 2 Critical review of resources linking Gender-based Analysis Plus (GBA+) and evaluation

Author	Acronym	Year	Total
Office of the Auditor General	OAG	2009 2015	2
Standing Committee on Public Accounts	PACP	2010 2016	2
Royal Commission on the Status of Women	(Byrd)	1970	1
Standing Committee on the Status of Women	FEWO	2005/06 2011 2017	6
Status of Women Canada	SWC	1998 2000 2007 2008 2009 2016	6
Treasury Board Secretariat	TBS	1977 1981 1984 1991 1994 2000 2009 2016 2017	11

		2018	
Other Government Departments	OGDs		8
TOTAL			36

Inclusion criteria were specifically applied to the search and retrieval of these resources focusing on the linkages between GBA+ and evaluation. As such, exclusion criteria were strictly applied to omit the following resources:

- evaluation reports drawn as part of a 5-year sample were undertaken for each organization participating in the case study line of evidence. While not representative of the entire universe of evaluation reports available, limitations were established because of the actual level of effort and resources available as part of my research. Future empirical studies examining the intersection of GBA/+ and evaluation represent an interesting opportunity to broaden the evidence base in this area; and
- resources focusing on either GBA/+ or gender-responsive budgeting or audit. While related, it was also beyond the scope of this research and similarly represents opportunities for future academic investigation.

I developed a protocol to assist in undertaking the critical review. This tool highlighted dimensions such as focus, structure, methodology and suggested methods or approaches, reasons/evidence, consultation, conclusions, and a critique of the underlying logic that were believed to be of significance in probing the intersections between GBA/+ and the evaluation function. Each of these resources was reviewed against this protocol. Although detailed findings are presented in subsequent thesis chapters; little evidence regarding the intersection of GBA/+ and evaluation was concretely noted until the release of the OAG's report on GBA in 2009. At this point, linkages become more prevalent.

Case studies

Case studies provide an excellent opportunity for the application of not only qualitative but quantitative methods (Thomas, 2011). The purpose of a case study lies in its specificity in unpacking “one instance” (Babbie & Benenquist, 2002). Stake (2000) elaborates on what comprises the one instance with his identification of three case study purposes that include intrinsic, instrumental, and collective rationales (2000, pp.437-438). It is the latter two purposes that apply to this research. They include the following: for instrumental instances in which the case is examined to mainly provide insight into an issue or to revise a generalization, and for instrumental instances where several cases are studied to investigate some general phenomena. My research employed the case study line of evidence to meet both of these purposes which are informed by the tailored ToC for integrating GBA+ in the federal government evaluation function. Case studies as a distinct line of evidence must consider the issue of generalizability (Silverman, 2018); some confidence exists not only because of the number of case studies completed but also because of the varying foci of each case study studied under the general phenomena umbrella. This necessarily involves making generalizations akin to “statistical inferences without employing probability criteria” (Silverman, 2018, p.268).

Originally, I anticipated onboarding between two and three organizations to participate in the case studies. The case study unit of analysis was envisioned to explore how evaluation units approached incorporating GBA/+ within one evaluation. This number was arrived at before the release of the 2016 *Policy on Results* which specifically articulated GBA as a federal government priority for the evaluation function. I purposively sampled organizations based on my perception of existing varying competencies to integrate GBA+ in evaluation. I had originally estimated 10 hours per month on average per organization based on my desired level of involvement. Original outreach communications were forwarded to 8 departments of varying sizes in terms of full-time

equivalents (FTEs) and resources, across various government sectors. These departments included:

- Global Affairs Canada (GAC);
- Department of Canadian Heritage (PCH);
- National Research Council of Canada (NRC);
- Social Sciences and Humanities Research Council of Canada (SSHRC) who shares a joint evaluation unit with the Natural Sciences and Engineering Research Council of Canada (NSERC);
- the Canadian Institutes of Health Research (CIHR);
- Economic and Social Development Canada (ESDC);
- Citizenship and Immigration Canada (CIC) now Immigration, Refugees and Citizenship Canada (IRCC); and
- Fisheries and Oceans Canada (DFO).

I made contact with these organizations because of previously established professional relationships, or because of an interest expressly articulated by the HoE, or Head of Performance Measurement (HoPM) who were aware of my research being undertaken. In-person interviews and/or electronic communications were undertaken with either the respective HoE or HoPM within each organization to determine the fit-for-research purpose and determine their alignment with my available resources. Ultimately, 6 organizations elected to participate in the research with some attrition expected during research.

Unequivocally, organizational research participation objectives highlighted requirements to build related capacities within evaluation units not originally envisioned as part of the research objective. How this developing capacity was envisioned to occur varied across participating

organizations. In hindsight, capacity building efforts related to these case studies could be identified as either fragmentary or systematic. Fragmentary efforts included my critical reflection during evaluation scoping efforts to identify how and what dimension of GBA to include in evaluations and at what level, the development of tools for use in scoping associated dimensions such as need, risk, costs, and level of effort necessary for consideration during evaluation efforts, provision of suggested feedback on evaluation data collection tools including interview guide and survey questions, as well as the provision of feedback on related evaluation reporting dimensions. More systematic efforts included the identification of specific outcomes to be delivered through intentional development of workshops held over 8 months with the focus of each one on a different evaluation phase. Systemic approaches to developing capacity included the development and delivery of focus groups/knowledge exchange activities to coordinate amongst various stakeholders, a common understanding of how GBA/+ might be integrated not only in evaluation but across functions involved in organizational results domains. Capacity building in light of the changing GBA/+ policy environment was one of the incentives driving organizational involvement. Organizations were also very interested in knowing what actions were being undertaken by others in this regard across the federal government evaluation function.

In three of the case studies, I followed a single evaluation over an approximate twelve-month period to determine how the organization approached integrating GBA/+ in their one evaluation. This duration was necessary to ensure that all elements of the evaluation lifecycle (planning/design, conduct, analysis, and reporting) were captured during data collection. Another organization had agreed to participate after careful consideration had concluded they would not be including GBA/+ in their evaluation. This organization bears special discussion given that it

represents what could be considered the “deviant” example in case study research (Stake, 2000). With the focus of my research on integrating GBA/+ in evaluation, it was important to examine instances in which this was *not* occurring given the newly identified evaluation priorities previously mentioned. The engagement of this organization in particular represented a significant learning opportunity given the intentional decision made at the onset to not include GBA/+ in their evaluation efforts. Opportunities exist particularly concerning the identification of challenges, and how organizations elect to address these challenges. The remaining two organizations intentionally focused on building capacities on integrating GBA/+ analysis in their respective evaluation functions, as well as within the organization’s function including the entire results domain. The process to onboard organizations in some instances lasted approximately six months in negotiating participation objectives, processes, and products. Overall, following six organizations with unique approaches to integrating GBA/+ in the respective evaluation units resulted in my original resource estimates being vastly underestimated.

Data collection efforts across case studies included at a minimum: a literature review specific to the focus of that evaluation or where evaluation function capacity was highlighted – the sector; a review of published evaluation reports over 5 years to highlight the state of the GBA/+ integration, a series of key informant interviews that included evaluation, performance measurement, program management, and gender expertise personnel where the positions existed. Unique to the two organizations electing to focus on capacity building, either a series of workshops (n=6) or focus groups (n=3) with both evaluation personnel as well as results division personnel occurred. The importance of this is discussed further in later chapters. Organization participation was caveated with the proviso that organizational findings would not be published as part of the dissertation, but rather, findings would be integrated across dissertation chapters.

Where produced, case study outputs (including reports and presentations) were provided to organizations as a value-added incentive to participate in research and were in some instances, very time-consuming to produce and finalize.

Case study data collection efforts were initiated in March 2017, and concluded in December 2018. This represented a 20-month data collection marathon; time spent on and with each organization varied from at a minimum of six months through to the full timeframe noted. Evidence for each case study was comprised of summary discussion points from electronic or phone communications, meeting minutes, organizational documentation, key informants' transcripts, focus group/workshop transcripts, reviews of previous evaluations, and relevant findings from the academic literature. In addition, one other organization was provided with a point-form outline using the same format, for the same reason noted above. In June of 2018, all case study organizations in addition to two other federal organizations (including Atlantic Canada Opportunities Agency [ACOA) and Canada's School for Public Service [CSPS} were invited to a case study debrief held at the University of Ottawa to share their efforts, identify emerging promising practices, and continue the discussion surrounding challenges and possible mitigation strategies. Minutes from this session were distributed to participants afterward.

Key informant interviews

Interview as a distinct form of data collection provides opportunities to engage with individuals to garner rich, detailed, and nuanced data (Patton, 2008) in direct comparison to survey research which may be regarded as insufficiently in-depth and may be superficial in their review of causation (Babbie, 1995). In the words of Brinkman and Kvale, interviews also allow for knowledge construction "in the interaction between the interviewer and interviewee" (2015, p.4) which is relevant to my research given the nascency of the integration of GBA/+ in the

evaluation function. Interviews for this research were semi-structured and incorporated a series of questions that were framed by both the outcomes, indicators, and assumptions identified in the tailored ToC. Key informant interviews also provided non-prescriptive opportunities to provide input and/or ask questions on the subject.

Key informant interviews comprised a distinct line of evidence for my research; distinct groups of stakeholders I had identified included:

- academics studying the integration of GBA/+ in the federal government domain;
- academics studying the integration of gender and other elements of diversity in evaluation processes and protocols;
- academics studying intersectional feminist analysis;
- evaluation practitioners with more than 10 years of experience in integrating GBA/+ and other elements of diversity in evaluation products and processes; and
- federal government employees from SWC and TBS responsible for the development and implementation of GBA/+ in the federal government evaluation function.

In total, 16 key informant interviews were undertaken to support this research. This number is independent of key informant interviews undertaken as part of the case study line of evidence. Minimal duplication (n=2) between the case study and key informant interviews was found. The total number was higher than the anticipated 10 key informant interviews originally proposed. The total number of key informant interviews was relatively low overall because of the relative nascency, and the limited expertise available both domestically and internationally.

All key informant sampling were non-probabilistic; I had focused instead on the purposive and intentional engagement of knowledgeable individuals. Interviewees were recruited through one of three mechanisms that included identification and contact by myself, recruitment after

completing a survey for the HoE, and referral chain sampling. For the first mechanism, individuals with known expertise in gender, evaluation, intersectionality, and GBA/+ analysis were identified. Sampling in this manner addresses critical cases, extremes, intensities, political importance, criterion, and typical case sampling types (Creswell & Poth, 2018). Admittedly this process was biased given my positionality previously discussed, however, this bias is reasonable given limitations in related expertise available. Attempts to mitigate this bias were reflected in the latter two mechanisms employed in recruiting key informants. Potential key informants were asked to self-identify after completing the HoE survey; five key informants were interviewed as a result of this line of recruitment. Additionally, chain referral or snowball sampling was also employed where at the interview's conclusion key informants were queried as to whether they believed anyone else should be interviewed for this research. Recommendations for possible key informant interviews were made in four instances. Three out of the four recommendations were accommodated. The final list of key informants, as well as the key informant interview guide, is contained in the appendices of this dissertation.

Interviews were semi-structured; meaning that many predetermined questions were identified but room for elaboration, and/or different foci were also permitted. This was important as it allowed space to address the rapidly changing context of GBA/+ implementation, not only broadly in the federal government, but in the evaluation function, to emerge. Once interview logistics had been confirmed, each key informant was forwarded consent documentation, and an interview guide to ensure that an ability to prepare if interviewees desired. After reviewing the interview, two interviewees (n=2) believed that they were not able to respond and declined to participate in the interview process. Interviews lasted approximately one hour on average and were conducted in person (n=4), by telephone (n=8), and by Skype (n=4). Generally, interviews

with those outside of the NCR were undertaken by telephone or by Skype. All interviews were recorded and later transcribed for analysis purposes although, in some instances (n=2), the poor quality of the recording meant that a full verbatim account of the interview was not available. Related findings are presented in the following thesis chapters.

Survey of the Heads of Evaluation (HoE)

A survey of the HoE was featured as a key line of evidence in my research. The HoE are responsible for the implementation of the federal *Policy on Results* within their department or agency in establishing an evaluation unit, and subsequent practice that meets their organization's needs in providing information to support decision-making. These individuals are appointed by their organizations' Deputy Heads (Treasury Board Secretariat, 2005) and are recognized as part of the organization's senior management cadre. Heads of Evaluation reporting requirements directly engage Deputy Heads (DH), organizational departmental evaluation committees, and senior management (Treasury Board Secretariat, 2013). As such, HoE shoulder evaluation policy and practice responsibilities at both the policy and practice levels within their federal organizations. In this role, they are well-positioned to respond to how GBA/+ is integrated in the federal government's evaluation function. I was intentional with purposive sampling (Schaeffer, Ott & Mendellhall, 1986) for this line of evidence because the respondents would be able to address survey questions based on their leadership position within organizations, and would additionally be able to address the extent of integration to date.

I designed the survey to reflect most dimensions identified in the tailored ToC. I had included 59 questions grouped into thematic areas regarding the integration of GBA/+ in evaluation including awareness, understanding, perceptions, supports, tools, capacity as well as utilization. The final section of the survey included logistical information that would assist in

disaggregated analysis and presentation of findings (organizational size and sector. This survey was comprised of predominantly binary responses accompanied by several open-ended responses. The closed-ended binary responses were comprised of yes/no options rather than scaled Likert responses to reduce mediating options available to respondents. I was intentional in this regard; my intentionality was predicated on three factors that included:

- The changing dynamic of GBA/+ within the federal government policy environment. particularly with the introduction of new policy requirements. It was important to document the *presence*, rather than the *extent* of integration efforts given the relative nascency amidst the evolving policy context. Further GBA+ evolutions were prophesized by federal government leadership and were reflected in Standing Committee on the Status of Women discussions tabled in Canada's Parliament;
- The time required by responding HoE. These individuals are very busy with little time to participate in academic research activities. While avoidance of nominal respondent choices is generally advised, a primarily binary scale was demonstrated to have been effective in reducing the response burden on survey participants (DeVellis, 2003). The consequences of approaching the HoE survey in this manner limited multivariate statistical analysis. Minimal open-ended response options were provided; and
- The format of this survey was informed by Skinner's work (2007) on measuring knowledge exchange outcomes. The application of knowledge outcomes at various stages including dissemination/diffusion, adoption, reach, uptake, utilization, transfer, and translation. This usage scale stems from Knott and Wildavsky's (1980) 7 categories of knowledge utilization and ultimately aligned with identified objectives of both GBA/+ and evaluation in their provision of information to support decision-making.

As previously discussed, as a result of professional experiences during my tenure in the federal government evaluation context many HoE and their contact coordinates are familiar. Access to these coordinates served as the basis for developing a current list of electronic contact coordinates to use for survey invitations. The remaining electronic coordinates were sourced from publicly available government employee directories such as the Government Employee Directory (GEDS), GOV411 Employee Directory, or LinkedIn. I compiled an initial list of 76 individuals, their organizations, and their contact coordinates in the spring of 2018. This list was uploaded to the SimpleSurvey platform (the software selected for survey administration) which houses secure databases on Canadian territory, to automate survey invitations, and reminders throughout the fall of 2018.

Best practices regarding administration of online surveys (Dillman, Smyth, and Christian, 2014) were observed including clear, concise questions, question grouping on common elements, early morning administration of survey invitations, and reminders administered early in the week. The HoE survey was available for respondents to complete in either official language; all respondents (n=29) completed the survey in English. The survey remained open from September to November 2017. The initial survey invitation was delivered late in September, and three reminders were issued over the next month and a half at weekly intervals. Table 3 Heads of Evaluation Survey Outreach Efforts presented below identifies outreach communication dates, and the total number of respondents having initiated the survey.

Table 3 Heads of Evaluation (HoE) survey outreach

Communication Date	Type	Emails/ Phone Call	Number of Surveys Completed
September 17, 2018	Invitation	76	7
September 25, 2018	Group Reminder	67	+3
October 2, 2018	Group Reminder	65	+4

October 9, 2018	Individual Email	62	+4
October 24,2018	Individual Email	59	
October 25, 2018	Phone Call	13	+5
October 26, 2018	Phone Call	11	+10

Several original invitation emails bounced (n=3). Additionally, I found that four (n=4) individuals had transitioned to another organization. Alternative contact coordinates were found for several (n=3) of the HoE and invitations and/or reminders were forwarded to these new coordinates. This survey had been delayed several times in wanting to be sensitive to the HoE work cycles, as well as being sensitive to the potential maturation of the linkages between the two functions. Response rates were quite low until I made direct phone calls at which point the total response rate had increased to the point at which it was comparable to others who had undertaken similar efforts (Fleischer et Christie, 2009; Galport et Azzam, 2016; Morris et Clark, 2013). SimpleSurvey's platform statistics noted the response rate at almost 33 percent and a completion rate of 79 percent.

During my phone calls with HoE, factors affecting their decision to participate in the survey including questioning the relevance of integrating GBA/+ in evaluation (n=1); pressed responsibilities related to their assuming multiple roles within their organizations including Chief Audit Executive (CAE) (n=3); a reluctance to participate because they felt it would reflect badly on their organization (n=4) as discussions of related efforts were just being recognized and action plans and/or training being undertaken; and, they were away on long term travel assignments (n=1) were identified. These explanations would account for some of the respondents who had initiated their survey responses but had not fully completed them particularly as the latter sections of the survey identified tools, supports, and use which would not be readily apparent in early maturation stages. Several HoE (n=2) had identified uncertainty as to whether they should

complete the survey given that they had participated in the case study line of evidence. A copy of the HoE survey can be found at the appendices of this dissertation.

Analysis and reporting

In the following thesis chapters, findings are presented with triangulation in mind. Principles underpinning triangulation regard validation as important but also reference purposes including enriching understanding, refuting confirming, and explaining as well (Carvalho & White, 1997). Multiple forms of triangulation exist; this extends to methods, to data, to the investigator, to theory (Denzin, 1973). With the multiple lines of evidence employed in my research, I believe that methodological triangulation is sufficiently addressed. Less onus was provided on investigator triangulation as I was the only researcher involved; this may present challenges for others who do not possess the same positionality that I do. Triangulation goals ultimately speak to addressing bias; I was a researcher aware of this from the beginning and attempted to “check my bias at the door” in a manner of speaking. This is referenced because often as researchers, much of the analysis that occurs, happens in the researcher’s mind during data collection (Patton, 2008).

Chapter 4 – Dear Betty: I have completed my mise-en-place. Now, what do I do?

This thesis chapter brings together findings from all lines of evidence regarding the enabling environment including a political and bureaucratic will, policy supports, tools, and training across the various stages of the tailored ToC. Additionally, awareness, recognition, and related perceived value and benefits are examined in greater detail in this thesis chapter.

Effective gender mainstreaming initiatives have identified several essential dimensions requiring inclusion for successful implementation (Standing Committee on the Status of Women, 2005, Parliament of Canada, 2005, Standing Committee on the Status of Women, 2016) that are discussed in this chapter. The Canadian federal government’s GBA+ framework maps core dimensions as including:

- A responsibility centre, to oversee the implementation of GBA+ and provide internal advice;
- An organizational needs-assessment, to determine what capacity and resources already exist in the organization, and to inform the creation of a work plan;
- A policy statement, or statement of intent, to articulate the commitment to GBA+ and provide a mandate for implementation;
- GBA+ training and tools, to facilitate buy-in, build capacity, and inform different parts of the organization about GBA+;
- A GBA+ “pilot” initiative, to provide a concrete example of GBA+ application; and
- Ongoing monitoring of progress, to highlight successes, best practices, and to identify gaps and new priorities.

All of these dimensions are discussed in greater detail in various sections of this dissertation; in the chapter Enabling Environment, Tools and Supports three of these framework dimensions

(responsibility centres including GCs, GFPs, policy statements, and training/tools) are discussed in relation to evaluation. The monitoring dimension is discussed in the following Risks and Mitigation Strategies chapter. The 2 remaining GBA+ framework dimensions that reference piloting and an organizational needs assessment are embedded in reported case study findings throughout this, and the remaining chapters.

Political/bureaucratic leadership and will

Changing attitudes towards GBA/+ are largely attributed to the recent shift in political leadership; the role of the Prime Minister (PM) as a key public policy influencer must be considered when these priorities manifest within and across the bureaucracy. With the election of Justin Trudeau in 2015, one of the key challenges noted in evidence produced by the critical review addressing senior-most political and bureaucratic will has been tempered. Until this point, the gender equality and GBA/+ implementation threshold persistently hovering at the edge of the policy landscape had shifted firmly into application. Temporarily at least, this one challenge seems to have been waylaid although the evaluation function lags behind the policy and program functions as targeted producers and users of GBA/+ information.

Gender equality as a federal government priority has been communicated to all individuals comprising the Prime Minister's (PM) Cabinet. These are the individuals appointed to ministerial positions within federal organizations (currently numbering 35 Cabinet members including Justin Trudeau). Ministerial mandate letters outline commitments in support of the federal government's top twelve priorities (Privy Council Office of Canada, 2018) including gender equality and the need to demonstrate progress towards established gender equality indicators. Each mandate letter outlines leadership obligations to be achieved during their Cabinet tenure. Each ministerial mandate letter contains "gender" at least once in stating their leadership role,

and related responsibilities. It can safely be assumed consequently that gender at a minimum, if not GBA/+, is present on senior management's organizational radar. Additionally, progress towards ministerial mandate commitments has been made publicly available for the first time via the federal government's *Mandate Letter Tracker: Delivering Results for Canadians* which is an online tool displaying real-time progress against these commitments. It was noted during the thesis finalization stages occurring during spring-summer 2020 that the federal government's mandate tracker has been temporarily closed.

Relevant to this research, the Minister of Status of Women's mandate letter specifically notes that Minister Monsef should "work with the President of the Treasury Board to ensure that GBA+ is incorporated into the Departmental Results Framework (DRF), Departmental Plans (DP), and performance reporting" (Prime Minister's Office of Canada, 2017). While not referenced directly, the evaluation function falls within the performance reporting domain. Ironically, gender and not GBA/+, appears only once in the Treasury Board President's mandate letter (Prime Minister's Office of Canada, 2017); its reference is employed relative to gender parity in leadership positions. This non-reciprocal relationship as reflected in the mandate letters is peculiar given that both Mr. Brison and Ms. Philpott (respectively, then Minister and President of the organizations referenced) were directed to "take a leadership role to review policies to improve the use of evidence and data in program innovation and evaluation, more open data, and a more modern approach to comptrollership" (Prime Minister's Office of Canada, 2017). GBA/+ would have been well-aligned for identification in this specific ministerial commitment.

Less than 2 years after making public ministerial mandate letters and progress towards these commitments, gender equality and GBA/+ have permeated organizational leadership's performance discussions. Almost all HoE survey respondents indicated being conceptually aware

of GBA/+ (97 percent). This number drops only slightly to 93 percent when queried whether or not respondents are aware of existing national commitments to integrate GBA/+ in assessments of programs, policies, legislations, and initiatives. This finding highlights that the evaluation function is aware of GBA/+ and the requirement to consider not only gender but other expressions of personal identity. Although these commitments are nearly universally understood, these commitments have also been in place since Canada's affirmation of the *Beijing Declaration for Action* in 1995; and more formally since the OAG GBA audits (Office of the Auditor General, 2009; Office of the Auditor General, 2015). This signals other factors at play in the recognition, implementation, and application of GBA/+ within the federal government's evaluation function.

Supporting policy

The specific articulation of GBA as a federal government priority in the 2016 *Policy on Results* (Treasury Board Secretariat, 2016), and more detailed elaborations in the *Interim Guidance on the Policy on Results* (Treasury Board Secretariat, 2017) translated to an evaluation priority. GBA/+'s integration stems from a concrete recommendation made by the OAG in 2015 and was identified in SWC's *Action Plan on GBA+ 2016-2020* which states:

The Treasury Board of Canada Secretariat (the Secretariat) periodically undertakes in-depth reviews of Treasury Board policy instruments. The Secretariat will maintain its internal guidance that gender-based analysis (GBA) considerations should be taken into account, as appropriate, when developing and reviewing Treasury Board policy instruments. As evaluation is one means of assessing program initiatives, the Secretariat will also support Status of Women Canada (SWC) by providing input to the development of guidance for evaluators on GBA and by assisting the dissemination of completed guidance to departmental evaluation functions through the Secretariat's established channels within six months of the completion of the guidance. When developing Treasury Board policy instruments related to program evaluation, the Secretariat will examine options to help departments and agencies consider government-wide policy considerations, such as GBA, when assessing the performance of federal programs (SWC, 2016, Recommendation 1.61

While the federal government evaluation community is taking GBA/+ seriously, shared reciprocity within GBA/+ policies does not exist. “Evaluation” is independently referenced on six distinct occurrences throughout the most recent GBA+ Action Plan; comparative to the 2008-09 *Action Plan on GBA* which references evaluation only twice, and only then in the context as a specific juncture in the policy cycle. In former GBA+ action plans and strategy documents, evaluation is mentioned in the following instances that include an evaluation of the effectiveness of the GBA process itself (Status of Women Canada, 1995, p.19), and in specific sectors such as health and violence against women (VAW) to provide insights into the policy development, research, analysis, and training in these areas (Status of Women Canada, 1995, p.34).

Evidence from the critical review identified that the GBA policy statement dimension of the GBA capacity framework can be traced back to the *1995 Federal Plan for Gender Equality*. In this strategy document, all federal departments and agencies were called upon to implement “gender-based analysis by establishing an analytical process, or framework for assessing the potential differential impacts of legislation and policies on women and men” (OAG, 2009). This call was echoed in recommendations presented before the 2005 FEWO report that “that all federal departments and agencies develop a strategic framework for GBA and action plans for implementation” (Standing Committee on the Status of Women, 2005) and reiterated in the Expert Panel’s review on the accountability mechanisms for gender equality. The OAG’s investigations into GBA (2009, 2015) judged departmental responses as being lackluster at best. Subsequent investigative efforts report initiated by FEWO noted that only 29 organizations of a possible approximately 110 had developed GBA frameworks (Standing Committee on the Status of Women, 2016, p.4). These organizations account for almost 60 percent of the federal government’s spending, however (Office of the Auditor General, 2015, Standing Committee on

Status of Women, 2016). Despite repeated calls for framework development, the 2009 OAG GBA report noted that considerable variation existed across organizations (Office of the Auditor General, 2009). They also noted that many of the formerly advanced departments had “aged out” their relevant GBA training, policy statements, focal points, etc. This implies an ongoing need to refresh these items with an accompanying resource investment.

A GBA/+ policy statement was identified as including the following four components (Status of Women Canada, 2017):

- An organizational commitment to GBA+ as a mechanism to advance gender equality in Canada;
- A statement of need for GBA+ with emphasis on its relevance to the organization’s work/issues;
- A clear indication of the organization’s commitment to both implementing and sustaining the practice of GBA+; and
- A requirement that GBA+ be a standard function of a systematic, integrative, and evidence-based approach to developing and evaluating research, policies, programs, and legislation.

In revisiting Women and Gender Equality’s (WAGE, formerly SWC) GBA+ website while finalizing thesis content, the final component was noted to have been removed within the last year. It is hypothesized that another section added on the recognition of GBA+ as a competency was intended to replace all references to evaluation, however, the removal of evaluation specific references has occurred on multiple occasions consequently prompting more questions than providing answers.

Slightly less than half of the HoE survey respondents (47%) indicated a familiarity with individual organizational GBA/+ policies, initiatives, or guidance materials. This number is low, particularly when the organizational affiliation of respondents was referenced against those organizations individually identified in the FEWO 2016 GBA+ implementation report. In this accounting, organizations indicated having developed organization specific GBA/+ materials. Of the 16 organizations in the HoE survey that identified not having specific organizational materials related to the GBA/+ frameworks, 5 actually do according to organizations cited in the 2016 FEWO report. This discrepancy suggests that the GBA/+ tools have not been widely circulated or promoted within their organizations given that evaluation functions are currently unaware of them. It further suggests several directed actions for future consideration including direct engagement with the evaluation function. Several case study key informants (n=2) with gender expertise identified that they had not yet had the opportunity to work directly with the evaluation functions within their own organizations. The evaluation function was cited as a secondary priority after policy and program areas were targeted for initial GBA/+ capacity building efforts. This is particularly relevant when considering findings from a question posed in the HoE survey that asked whether it was believed that the evaluation function within their organization would be affected by these policies, initiatives, or related guidance materials. Ninety-two percent of respondents identified that the evaluation function would be affected by organizational GBA/+ frameworks.

Despite early gender equality frameworks calling for the “acceleration of gender-based analysis” in the *Agenda for Gender Equality* (2000-2005), a difficult to access evaluation of three dimensions of the AGE strategy obtained through Library and Archives Canada highlighted significant challenges to the federal government’s approach to gender equality and GBA. The

reports self-identified as a challenging report and positioned the gender equality and GBA agenda as an ad-hoc/piecemeal initiative. The evaluation reported that the magnitude of the gender equality objective was incommensurate with the resources invested (Status of Women Canada, 2005, p.7). Simply put, AGE was inadequately funded, insufficiently staffed, and had little capacity. The following recommendations were made to support future widespread implementation of GBA that included (Status of Women Canada, 2005, pp.iv-v):

- The requirement for stronger and clearly articulated support from political and bureaucratic sources at the highest levels. Leadership needs to be more pro-actively supported by genuine working equivalence with senior-most levels within the federal public service;
- A legislative/regulatory framework that compels reporting across the federal spectrum on progress and which can lead to remedial action as called for;
- A need for a fully developed, completely documented conceptual framework and operating plan for all aspects of the AGE; and
- A need for more, and more effective and regular consultations and collaborations with groups seeking equality for women across Canada.

Key informants expressed a range of familiarities with GBA/+ which is a uniquely Canadian adoption of intersectionality to support gender mainstreaming efforts. Canadian key informants were intimately familiar with GBA/+. International key informants also consistently identified an awareness of integrating not just gender, but other dimensions of personal identity in evaluation products and processes. Some variance existed amongst interviewees regarding what personal identity elements should serve as the entry point for analysis; they drew on personal experiences to highlight where according attention to various elements of personal

identity beyond gender or sex, such as race or socio-economic status had impacted the results of their research or evaluation.

While the *Policy on Results* (Treasury Board Secretariat, 2016) plays an important role in determining evaluation scope, focus, and implementation, multiple lines of evidence from this research also point to the role played by the introduction and emphasis of GBA/+ in other policy cycle junctures as well (highlighted in the second thesis chapter). Government key informants and case study participants highlighted with equal importance, the stricter perusal of GBA+ requirements in Budget submissions, TB Subs, and MCs as key drivers in this integration process affecting those in the results domain. Evaluation personnel are often invited to participate in development discussions or communications in the TB Sub and MC proposal development process. During development, evaluators may be tasked with providing methodological advice in establishing an intervention's outcomes, indicators, and subsequent data collection efforts but program representatives are considered as functional leads. Consequently, GBA+ may not be the priority during these discussions but they can certainly be raised by evaluators for consideration. Further, an opportunity exists during these discussions to raise related problematic concerns regarding GBA+ information. For example, if GBA/+ information is not currently collected, a plan may be developed as part of the TB Sub or MC development process to ensure that funding is available to support such efforts which would ultimately facilitate access to information required to integrate GBA/+ in subsequent evaluation efforts. Interestingly, several case study key informants indicated that within their organizations GC, nor GFP were not invited to contribute to these discussions.

Relevance, value, and benefits of integrating Gender-based Analysis Plus (GBA+)

Because of GBA+'s historical associations with feminist movements, it was originally questioned as to whether stakeholders associated with the federal government's evaluation function with its focus on neutrality and objectivity would be able to identify GBA's relevance to, and within evaluation efforts. Respondents to the HoE survey affirmed at 88% that it was relevant to include GBA/+ in their evaluation efforts. Exploring this association in further detail, only two respondents correctly identified the relationship between GBA+ and feminist knowledge creation. The remaining respondents (n=23) reported that GBA+ was not considered to be "feminist" in their organizations. When queried further and in response to probing questions exploring whether "feminist" carried a connotation, only one response to each of the possible responses (yes/no) was provided. That only one respondent responded to each option is telling in itself. Additionally, when asked whether this connotative was positive or negative, the open-ended response provided was "not sure".

Key informant interviews clarified that speaking authoritatively regarding relevance was facilitated when government policies, programs, processes, and interventions directly affected individuals. This focus on the individual level ensured a much clearer "sight line" to *impact* and associated impact metrics. This notion is reinforced by early targeted capacity building efforts with socio-economic interventions (Standing Committee on the Status of Women, 2005) in which gender equality indicators variances were noted in such areas as employment, health, and leadership. This application to low-hanging GBA/+ fruit, or easily inferred inequality differences, was emphasized in the critical review evidence (Standing Committee on Status of Women, 2005, Standing Committee on Public Accounts, 2016). This holds particularly true when the *outcomes* of policies, programs, processes, and interventions refer to or provide delivery or services directly to individuals.

The difference between early-stage outcomes and impact is vast. Differences exist in both time (number of years), and the scope of the change required. In the results domain, impact occurs long after an intervention has ended whereas outcomes speak to what occurs while policies, programs, processes, and interventions are still, generally under government purview. Examinations of immediate, and intermediate level outcomes commonly are addressed by evaluation functions in their evaluations of government interventions which occur every five years; impact evaluations occur with less frequency. This latter element is discussed in some detail later in this thesis. The federal government evaluation function then would focus on early outcome levels with less attention accorded to the impact level in a results chain. This nomenclature (impact/outcomes) is imbued with specific meanings employed domestically, and internationally for various evaluation functions. Interestingly, a specific definition for impact isn't provided in TBS's *Results Based Management Lexicon* (Treasury Board Secretariat, 2015). Arguably, impact information would likely hold greater appeal for senior decision-makers, but reporting support at earlier outcome levels is required. The deployment and emphasis on *impact* by the federal government's gendered expertise may have actually served as a barrier to the implementation of GBA/+ by the federal government evaluation function.

The issue of integrating GBA/+ in reporting at impact levels was raised by key informants, as well as within several case studies that are specific to the federal government evaluation function. Two related issues were raised that included crossing departmental mandates, and the indicators employed. Simply, to order to achieve societal change reflected at impact levels, it is unlikely that accomplishment could be undertaken by one federal government organization with one policy, program, process, or intervention. Instead, many contributions from various actors within the federal landscape would feed into evidenced changes at this impact level. Several key

informants spoke of perceived “fiefdoms” in which organizations were viewed as acting protectively of initiatives (and related funding) under their purview. This raises an interesting debate about how cross-jurisdictional mandates may require collaboration amongst various federal organization’s stakeholders at a broader, and one might argue, more strategic level. Suggestions related to the crossing of organizational mandates to meet the needs of those experiencing inequality are met with challenges. These challenges may include altering central agency processes, existing terms, and conditions (Ts&Cs). This extension of a typical 5-year program renewal period may permit impact level changes to be realized. In addition, standardized indicators rarely serve as reporting metrics of individual initiatives; these are reserved for impact level changes and may include gender equality standardized indicators such as those identified under the Gender Results Framework (GRF) associated with recent Budgets. At the individual level, and especially for those programs, policies, processes, and interventions that were not subject to new GBA/+ requirements, unless one of these initiatives would have had a specific element of diversity embedded in the program, other metrics would have been more important to senior decision-makers.

While federal government evaluators may rely strictly on adherence to evaluation policy requirements; the refrain “because the policy says so” was common across case study organizations with developing capacities. As capacities evolved, so too did the relevance provisions made. Evaluation personnel were able to identify the GBA/+ conceptual issue which is discussed in greater detail in later thesis chapters. Stepping back from using the individual as the unit of analysis, provided an interruption to the perpetuation of the status quo of programs, policies, processes, and interventions being gender-neutral. One of the important components of GBA/+ is questioning the assumption “Is it really gender neutral”? To this point, a firm “no” was

recently recommended by FEWO (Standing Committee on the Status of Women, 2017). Central agencies who provide the challenge function will be directed to return any “gender neutral” or “GBA+ not applicable” proposals to forwarding organizations for improved and positive responses. Determining relevance by evaluators was raised by several key informant government interviewees as this responsibility falls to program management. During this transition to the new *Policy on Results* (Treasury Board Secretariat, 2016), it was noted by evaluators that often the program was unaware of their GBA/+ responsibilities having not been subject historically to new GBA/+ requirements. Evaluators then bore responsibility for explaining to program management that the program would be subject to some inclusion of GBA/+ when the program may not have been designed to account for any potential diversity elements. Key informant, and case study evidence found that preparation helped be able to respond to relevance questions posed by program management. Advance preparations included reviews of occurrences in other jurisdictions, and of both the academic and grey literature.

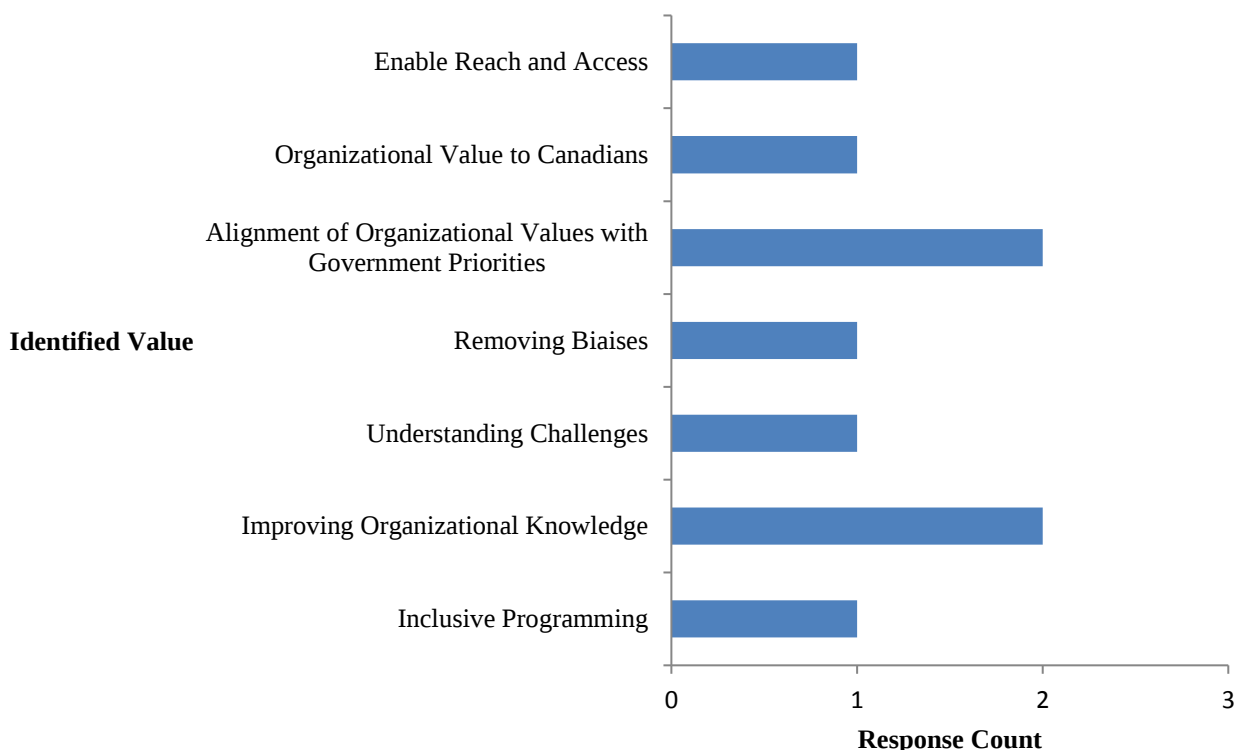
Regardless of GBA/+’s historical relationship with “feminism” and perceptions of relevance, the value, and benefit of integrating GBA/+ and evaluation is quite clear to the federal government evaluation community. Seven publicly declared value statements regarding the potential values of GBA+ have been articulated (Status of Women Canada, 2017):

- Consider the whole person, including multiple aspects of identity not limited to gender;
- Clarify diversity in people’s lives based on factors such as gender, sex, ethnicity, age, class, geography, and sexual orientation;
- Generate crucial information about the range of needs, priorities, capacities, experiences, interests, and views of differently situated women, men, and gender-diverse people;

- Identify barriers and negative impacts of policies, programs, and legislation related to gender and other aspects of identity;
- Allow government policy and program design to take into account the various factors that impact diverse groups of women, men, and gender-diverse people;
- Ensure policies and programs apply to a more significant portion of the population; and
- Recognize and accommodate human rights and support all citizens in fully realizing their potential as members of Canadian society.

These potential value statements target both the needs of individuals and public administration. Four public administration value statements exist that include: the provision of information, barrier identification and negative impacts, design responsiveness, and broader reach to diverse individuals. Seventy-seven percent of respondents of the HoE survey identified that they were able to identify the value of integrating GBA/+ in evaluation efforts. While promising, this represents a significant drop from levels identified with conceptual awareness and principles of GBA/+. In an open-ended response question, HoE were asked what these values were specific to the federal government evaluation function. Unsurprisingly, of the 18 who had responded their responses flagged public administration-related specific values. Figure 5 below notes these values and the response count.

Figure 5 Identified values of integrating Gender-based Analysis Plus (GBA+) in evaluation



It is noteworthy that respondents (n=18) had identified public administration values consistent with those previously identified. Additionally, respondents also identified a previously unreferenced value “alignment of organizational values with government priorities”. This was referenced more than once (n=2). Respondents were also queried about the perceived benefits of applying GBA/+ in evaluation efforts. Similar responses were repeated but there was also an emphasis on conditions uniquely aligned with the federal government’s evaluation function that speaks to the assessment dimension of GBA/+. These unique dimensions included examinations of impact, and effectiveness including relevance, performance, and efficiency. Figure 6 below presents identified benefits of integrating GBA/+ in evaluation efforts by response count.

Figure 6 Identified benefits of Integrating Gender-based Analysis Plus (GBA+) in evaluation



Responses indicated that some respondents found it difficult to distinguish between the perceived values and actual benefits of integrating GBA/+ in evaluation. This insight is offered based on three instances in which similar responses to the previous question were provided (knowledge and insight, alignment with government priorities, and reach and engagement). Failing to distinguish between perceived values and actual benefits may be attributable to delays encountered in case studies associated with a single evaluation report. In these instances, insufficient time would have elapsed to realize actual benefits associated with integrating GBA/+ in evaluation efforts. These delays impeded following a federal government evaluation from the planning stage to report dissemination. Of the single evaluation case studies, only two were able to follow evaluation efforts comprehensively from start to finish. The remaining two case studies focused on capacity building. In the single evaluation case studies, two, therefore, were not followed from tabling the evaluation report before the organizations' Performance Measurement

and Evaluation Committee (PMEC), and the development of a Management Response and Action Plan (MRAP).

Evaluation and Gender-based Analysis Plus (GBA+) responsibility centres, Gender Champions (GCs), and Gender Focal Points (GFPs)

Evidence presented by academics tabled in the 2005 FEWO report spoke to the need for leadership to champion GBA. A direct quote from Pauline Rankin references this requirement “[t]o be successful, gender-based analysis needs strong champions. It needs strong champions at the political level and it needs strong champions within the highest level of Canadian bureaucracy. I think those champions are pivotal to whether or not gender-based analysis works” (Standing Committee on the Status of Women, 2005, p.50). Gender Champions have been described as being responsible for (Standing Committee on the Status of Women, 2016, p.34):

- promoting awareness of GBA+;
- supporting training and GBA activities; and
- communicating the importance of GBA+ within the organization.

Gender Champions are thus considered as their organizations’ expertise on GBA+ (Standing Committee Status of Women, 2016, p.34). Gender Champions until 2016 met regularly once each year; however, meetings are now scheduled twice a year.

Recommendations tabled in FEWO’S 2005 report were addressed in the Government’s response that same year. Central agencies including the PCO, TBS, and the Department of Finance (DOF) were called to elect senior officials to the GC’s role within their respective organizations to support both internal gender equality, and GBA/+ oversight responsibilities. This included developing capacities to support their organizations’ challenge function roles. Additionally, departments and agencies were “encouraged to put in place gender-based analysis mechanisms and name champions as appropriate to each organization’s culture and mandate and

be prepared to report on achievements as soon as possible” (The Expert Panel on Accountability Mechanisms for Gender Equality, 2005).

The monitoring of GBA/+has been met with some degree of cynicism amongst federal organizations. Concerns were articulated regarding the appropriate level of seniority, and placement of that individual within an organization. This directly led to Recommendation 11 being tabled in FEWO’s 2016 GBA+ implementation report (2016). Related content differentiates between political and bureaucratic support. This concern is prompted in identifying two distinct roles that include: an individual at the assistant, or associate deputy minister (ADM) level being responsible for promoting GBA/+; and, a separate individual being responsible for undertaking the role of GC (Standing Committee on the Status of Women, 2016, p.37).

Similarly, the federal government evaluation function has struggled with appropriate access to, and levels of responsibility assumed by senior decision-makers within their organizations. In a novel undertaking in the 2009 *Policy on Evaluation* (Treasury Board Secretariat, 2016), it was identified that the HoE should have “direct and unencumbered access to the deputy head, as required (Treasury Board Secretariat, 2009). In subsequent reviews of the health of the evaluation function (Treasury Board Secretariat, 2010; Treasury Board Secretariat, 2011; Treasury Board Secretariat, 2012), it was reported that unencumbered access to deputy heads had increased each year. Unencumbered access is interpreted to mean regularly scheduled bilateral, or impromptu bilateral meetings with their deputy heads (Treasury Board Secretariat, 2012). In the last year that a health of the evaluation function report was publicly released (Treasury Board Secretariat, 2012), all large departments or agencies (LDAs) had a Departmental Evaluation Committee (DEC). These committees were chaired by a deputy head 85 percent of the time

(Treasury Board Secretariat, 2012) an increase from previous years, and 64 percent of these committees served only the evaluation function.

Despite HoE survey respondents reporting that their organizations had internal access to GBA/+ expertise (44 percent) and current sufficient capacity to integrate GBA/+ in evaluation efforts (44 percent), only 29 percent had consulted with their organizations' GCs for assistance in integrating GBA+ in their evaluation efforts. This statistic suggests that the mandate of the GCs may not be best aligned for consultation with the evaluation function despite their acknowledged organizational expertise on GBA+. This is particularly significant in light of the increase reported when respondents were queried on whether they had consulted with their organization's GFPs. Almost 40 percent of respondents to the HoE survey noted that they had consulted with the GFPs when integrating GBA+ in their evaluation efforts. This suggests that the working level of GFPs is more aligned with evaluation processes and products. One interpretation of this level of engagement is that GFP's possess the ability to provide more useful information to the evaluation function when applying GBA+ to specific policies and programs. Alternative explanations also include that despite being recognized as an organizations' GBA+ expertise, these senior officials may be too busy to engage directly with evaluation personnel, especially if they hold both the GC position plus another. Respondents also identified that both internal, and external resources were engaged to support integrating GBA+ in evaluation. External resources included: other internal resources, GCPedia, TBS/TBS units, SWC/WAGE, the Canadian Evaluation Society (CES), Human Resources (HR), other departments, academia as well as the broader evaluation community and external consultants. Many government key informants identified that they were lacking the internal expertise within their respective

evaluation functions, and resources were required to build this capacity. Resources required to building capacity included time, expertise, and finances.

Case studies reflected a varied mix of evaluation's engagement with their organizations' GCs. This variation ranged from a strong GC presence which was reflected in organizational priorities to another organization in which no GC, nor EDI Champion was present as the position was either unstaffed or the GC had moved on and the position had yet to be filled (n=3). EDI Champions have been noted in the context of case study findings because organizations often have initiatives for both GBA/+ as well as EDI running concurrently. These initiatives have a range of relationships with EDI which vary in meaning from *EDI is GBA+*, *EDI has some relation* with GBA+ which may be clarified in direction, and *EDI has no* relationship with GBA+. In the two organizations deemed to have an "established" capacity to integrate GBA/+ in evaluation efforts (n=2), there was a very close relationship with EDI. It should be noted that instances in which EDI was GBA+, EDI focused on four specific dimensions of personal identity which included gender, visible minority status, disability, and Indigenous status). Participating case study organizations with advanced capacities in integrating GBA/+ in evaluation efforts, GC's played a visibly significant role in which their related responsibilities included:

- Sitting at and actively participating in senior executive decision-making discussion tables;
- Visible engagement in and reporting on gender-based analysis activities within the organization; and
- Actively promoting and conducting research on GBA within their organizational context.

Additionally, one key informant had identified that the GC position required a degree of independence from the political bureaucracy to maximize effectiveness. In this instance, the GC

was appointed from an academic position which was perceived to be beneficial to possess independence within the political bureaucracy.

Uptake of Gender-based Analysis Plus (GBA+) tools and training

One of the essential components of an organizational GBA+ framework is the development of relevant tools and training (Status of Women Canada, 2017). Training materials related to GBA can be traced back as far as 1998 (Status of Women Canada, 1998); this occurred shortly after Canada's commitment as Beijing signatories. In these original materials, 8 steps were identified in undertaking GBA which differs from the current 7 steps (Status of Women, 2018). The final step "evaluating the quality of the gender-based analysis" has been removed. The training was thought to "help transfer the practical knowledge required to implement GBA+ within an organization, and it prepares employees to apply GBA+ to their own work" (Status of Women Canada, 1998).

Initially, GBA training focused on capacity building efforts delivered through direct engagement with individuals in *policy and program* [emphasis added] areas in 24 departments (Standing Committee on the Status of Women, 2005, p.19). To meet the needs of these target audiences, a policy and program framework was used and case studies developed to "reflect how the socio-economic realities of women and men were different; how to question assumptions about gender roles; and how to integrate gender considerations in all the relevant steps of policy and program development" (Standing Committee on the Status of Women, 2005, p.19) including evaluation. The 2005 FEWO report also identified that individuals were willing to undergo training but implementing systematically, and monitoring this implementation within their organizations was believed to be beyond their scope. Training efforts officially began in 2001; examples were found to be necessary to include in training materials. This need for practical

examples has particular relevance for tools developed for and used by the federal government evaluation function.

Respondents to the HoE survey identified that they were 100 percent aware of the federal government's GBA+ online training tool housed on SWC's website. Interestingly, this number drops to a reported 69 percent when asked whether they have received any training for implementing GBA+. These statistics provide insight when examining whether evaluation functions across the federal government make this training mandatory. A key informant interviewee noted that "only 22% of departments make GBA+ training for evaluators mandatory and 41% recommend it; while 19% of departments do not require any training" (personal correspondence, March 16, 2018). These findings were drawn from the annual implementation survey on GBA+ for 2016-2017. The online training GBA+ modules are available free of charge and provide an introduction with specific examples from various sectors.

Specific guidance materials on integrating Gender-based Analysis Plus (GBA+) in evaluation

The online GBA+ training is not geared for evaluators specifically; many organizations including those participating in case studies (n=4) have identified engaging consultants to assist in building capacities in this area. Training material modification is consistent with accounts of tailoring GBA/+training materials for specific applications (The Expert Panel on Accountability Mechanisms for Gender Equality, 2005; Standing Committee on the Status of Women, 2005). Consequently, SWC's online GBA+ training materials may be helpful as a conceptual introduction but not as helpful in practical application for the federal government evaluation function. At this juncture, training materials specific to the performance measurement were not available, although one of the case study organizations noted that the HoPM had indicated moving in this direction.

Discussions with government key informant interviewees and case study participants highlighted specific training requirements relevant for their evaluation function that reflected their organizational context. This was explained as organizational mandates affect the type of evaluation that organizational evaluation functions undertake. Training must consider both the evaluation function and gendered expertise but in addition, must also reflect unique organizational attributes. These training materials must be methodologically specific across the entire evaluation lifecycle including planning/design, conduct, analysis, and reporting phases. Training must also address all potential qualitative and quantitative lines of evidence that could potentially be employed in evaluation efforts. Participating case study organizations that focused on building capacities spent a substantial amount of time discussing what integrating GBA+ might look like in specific evaluation data collection tools and what language might be used in surveys, key informant interview guides, focus group guides, etc. It was also expressed that training would help identify key evaluation questions and indicators addressing dimensions of GBA+. To date, specific training of federal government evaluators on integrating GBA/+ in evaluation has not been offered by any branch of the public service which has left the evaluation community struggling. With few training opportunities available in the private sector, and no training provided by the government, each organization has been left to its own devices. This refrain was echoed consistently across case studies and key informant interviews. The major stakeholders in this results domain have yet to offer training opportunities although professional associations (such as the Canadian Evaluation Society (CES) or the Planning and Performance Exchange (PPX)) have dedicated a portion of their 2018-19 learning program to GBA+ (National Capital Chapter Canadian Evaluation Society, 2019; Planning and Performance Exchange,

2019). This signals the importance of GBA/+ to the membership of these organizations who are comprised largely of federal public servants.

Several government key informants referenced the development of tools unique to their organization to assist their evaluation function in integrating GBA+. These tools were developed using two primary sources that included a draft version of related guidance materials, and content from SWC's online GBA+ training. Several of these tools are currently undergoing pilot testing. In another instance, a case study organization had spent several months standardizing their GBA+ questions for inclusion in evaluation matrixes (evaluation questions and indicators), key informant interviews, and focus group guides as well as survey instruments. The intention of such a resource-heavy investment was associated with the systematic capturing of GBA+ data for all future evaluations. These tools have been widely circulated throughout the federal government evaluation function through informal lunch and learn (L&L) sessions and informal networking opportunities. Sixty-seven percent of HoE survey respondents noted having access to tools to support organizational efforts to integrate GBA+. Further, 42 percent of respondents noted that their organization had developed unique tools to support integrating GBA+ in evaluation. If case study organizations are eliminated from this sample, another 8 organizations have independently developed unique tools. The number of organizations that have developed their own GBA+ integration tools for the evaluation function echoes the need identified above to reflect GBA+ in their organizational context.

Following the release of the 2009 GBA OAG audit, several government key informant interviewees identified that the federal government evaluation function could potentially fulfill the "stick" role in the carrots, sticks, and sermons policy typology (Bemelmans-Videc, Rist & Vedung, 2007). The stick role was believed to be appropriate given the synonymous assessment

dimension shared by the two functions and that as an established function, evaluation carried a greater weight for action by decision-makers with each evaluation requiring a management response. It was suggested after tabling of this audit, that related guidance materials be developed. At the end of the 2010 fiscal year, a small contract was awarded to support the development of a discussion draft for integrating GBA in evaluation. This initial draft development was overseen by an advisory committee and several iterations resulted before being shared with TBS. What resulted was not the envisioned product, and a limited release of a scaled-down version was promoted only at local evaluation events such as the 2011 Annual Learning Event (ALE) of the National Capital Chapter of the Canadian Evaluation Society (National Capital Chapter – Canadian Evaluation Society (NCC-CES)). Another concerted effort related to guidance materials was initiated in 2015 to coincide with recommendations flowing from the OAG's report tabled in that same year. However, the release of a new *Policy on Results* (Treasury Board Secretariat, 2016) and radical restructuring of the government's results environment including the evaluation function, caused additional delays.

Key informants flagged that one of the chief concerns was the lack of examples available for inclusion in draft guidance materials. To address these concerns, a phased approach was planned. Originally, it was anticipated that two releases of guidance materials were intended which included an introductory resource, and a more advanced resource. Where possible, these guidance materials would include domestic examples. Key informants identified that as GBA+ competencies were established within the Canadian federal government, an increased number of examples would be available. Early guidance materials did include examples, but many of these examples include those drawn from international settings and represented "low-hanging" GBA/+ fruit. These examples pose concerns for the evaluation function when critically reflected upon.

For example, in the introductory guidance materials, a transportation example related to snow removal in a Norwegian country was used. When investigating further, it was found that this GBA was undertaken not as part of an evaluation, but rather commissioned as a separate initiative. This has resource implications as the scope of the two efforts varies significantly.

Case study participants also flagged multiple concerns with the draft guidance materials released in a limited fashion to the federal government evaluation function in June of 2018. Identified concerns included: being consistent in differentiating between gendered and feminist approaches to evaluation, language employed (i.e. gender sensitive indicators, impacts/effects/outcomes), the inclusion of key GBA/+ dimensions such as the identification and reduction of barriers to support access and participation, and defining intersectionality. These are but a few of the most significant challenges facing TBS as they worked towards finalizing draft guidance materials. Eighty percent of the respondents to the HoE survey indicated a familiarity with the draft guidance materials on integrating GBA+ in evaluation but this percentage dropped to 75 percent when asked whether their organization's evaluation function had access to these draft guidance materials. Given that almost a full decade has passed since initiating efforts, it may be reasonable to suggest that the delay significantly impacted related capacity development relative to federal government evaluation functions. The delay in presenting guidance materials to the federal government evaluation community has however resulted in many organizations demonstrating initiative in developing tools unique to their environments. For example, the Public Health Agency of Canada (PHAC) and Health Canada (HC) who share an evaluation directorate, have developed a tool entitled the *SGBA+ Lens for Evaluation* and are currently in the process of developing supporting guidance materials for their joint evaluation function (National Capital Chapter-Canadian Evaluation Society(NCC-CES, 2019). They have modified

their approach from the GBA+ to include sex “S” as a key analytical variable given its importance in health and public health disciplines.

Other additional challenges interpreted as risks to GBA+ implementation efforts and possible mitigation strategies are discussed in greater detail in the following chapter of this dissertation.

Chapter 5 - Dear Betty: My oven is preheating. What temperature should it reach?

This thesis chapter explores the major risks regarding integrating GBA/+ into the Canadian federal government evaluation function. These risks can be classified both broadly at levels applicable to gender mainstreaming, as well as those risks that are more specific to the evaluation function. Specific risks include disaggregated data and clarification of levels appropriate for evaluation, the missing accountability stick, defining necessary and sufficient resources, sustaining momentum, at risk of becoming rote, quality of GBA+, and when interventions are grounded in elements related to GBA/+ (i.e. an initiative dedicated to multiculturalism, or disability). Where federal organizations have developed mitigation strategies to address these concerns, the relevant information has been included although readers should note that much like GBA/+ itself, what works in one organization may not be directly applied to another organization. The discussion of these risks is directly applicable to one of the assumptions forwarded in the tailored COM-B ToC for integrating GBA+ plus in the federal government evaluation function.

Access to disaggregated data by identity factor

One of the most pressing issues for integrating GBA/+ is the creation, access to (Standing Committee on the Status of Women, 2016), and use of disaggregated data. This concern affects the implementation of GBA/+ both as a gender mainstreaming strategy and the federal government evaluation function. Disaggregated identity data is one of the foundations of intersectional research as it potentially highlights status discrepancies at the individual level. Evidence from the critical review identified that while disaggregated data was referenced in the 2009 OAG's GBA report, this was not the case in the 2015 OAG report. More nuanced data languaging was employed in the later OAG report with references to "gender data" (Office of the

Auditor General, 2015); disaggregated data did not appear in the report. The return to disaggregated data features prominently in the FEWO 2016 account on implementing GBA+.

“Witnesses shared that the availability of relevant, reliable and complete data disaggregated by gender and other identity factors, as well as complementary analytical research, are crucial to performing GBA+” (Standing Committee on the Status of Women, 2016, p.40). It is also interesting to note that early inferences made regarding disaggregated data focused on quantitative data with multiple references to statistics as the main evidentiary support. However, as demonstrated with the last quotation we can see the entry of a qualitative data perspective with the alluding to “complementary analytical research”.

Significant resources have been invested in the creation of a Centre for Gender, Diversity and Inclusion Hub housed at Statistics Canada (StatsCan) which was launched in the fall of 2018. The purpose of this initiative is to provide quantitative data across various areas including the GRFs associated with budget exercises. The Hub’s consultation website directly associates quantitative information being available for “the evaluation of programs, policies, and initiatives from a gender, diversity and inclusion perspective” (Statistics Canada, 2018). This positioning of impact level quantitative data being available for evaluation purposes suggests a misaligned understanding of the level targeted by evaluations generally. Disaggregated data is required at multiple levels across a results chain, not only at impact levels. Directing the focus to impact levels leaves evaluation functions with the need to recreate disaggregated data at prior results chain levels to meet the needs of programs, policies, processes, and interventions that are unique to individual federal organizations. Several key informants referenced existing heteronormative biases in the collection and reporting at impact levels. For example, several federal government

organizations do not collect gender non-binary information from individuals as part of national statistical efforts.

The timing at which this collection of disaggregated data occurs was not introduced in the critical review. Optimally when evaluation efforts are initiated, the program would have years, perhaps even decades of GBA/+ information available to them; this however is often not the case given changes that occur as part of Ts&Cs renewal. In such instances, the evaluation team is then left with the responsibility of collecting this information themselves during the evaluation dependent upon to what extent and how GBA+ is addressed in the evaluation framework. Case study participants spoke in-depth on the challenges of providing disaggregated data for use by respective evaluation functions. In most instances (n=5), access in advance to specific evaluation level disaggregated data was lacking in part because of the time required, coordinated efforts, and compelling requirements to implement systematic data collection mechanisms. The exceptional case study participant with advanced related capacities had initiated efforts well before the 2016 *Policy on Results* (Treasury Board Secretariat, 2016) was implemented. Their demonstration of foresight towards integrating GBA/+ information requirements in their Performance Information Profiles (PIPs) as soon as the requirement was tabled as part of the new results policy environment. This involved information technology (IT) initiatives to amend applicant funding and beneficiary information to include GBA/+ related data.

These cross-organization efforts had required the collaborative engagement of multiple stakeholders including program, policy, evaluation, and performance measurement representatives as well as input from privacy, financial, and risk representatives. Significant financial investments were made in support of these organization-wide efforts. However, even in this instance, disaggregated data challenges were apparent as applicants were asked to self-

identify GBA/+ information. Self-declared information was perceived to be problematic because little opportunity existed to validate this information which could conceivably lead to flawed results reporting. Additionally, the collection of this information was undertaken in advance of funding being awarded. This supports the suggestion that timing may be an important consideration in the collection of GBA/+ plus information. Case study participants were inventive in their development of strategies to assemble disaggregated data for evaluation purposes: including merging existing program databases using multiple GBA+ variables, including narrative preambles introducing why and how GBA+ data were going to be used in specific evaluations, and other organizations being very specific about the languaging employed in data collection tools to rebut heteronormative assumptions. Participating case study organizations were also innovative in their approaches to recreating data at earlier results levels to support integrating GBA+ in their evaluation efforts. Two organizations had integrated data from existing databases within the departments to derive the gender and diversity information required. Limitations associated with this innovative approach are predominantly time-related although must federal organizations also must consider dedicating resources to cleaning/recoding data within a new database. Missing or incomplete data (such as dates) was also a concern that was articulated which is important as it provides a timeframe by which to examine initiative effects.

Similarly, key informants spoke to the need for, and efforts to create disaggregated data during evaluation efforts. Despite GBA+ being a tailored intersectional approach to gender mainstreaming, disaggregated data was uniformly identified as a minimum threshold requirement. Optimal conditions in which disaggregated data was available in advance of evaluation implementation were noted, and international stakeholders referenced policies in

places in organizations such as the United Nations family of organizations in which policies directing disaggregated data for every evaluation project are in place. Key informants working in international settings recalled instances in which disaggregated data was used to demonstrate to decision-makers including funders, evidenced gaps between certain groups of individuals. This information was used to ensure that subsequent decisions were made to enact changes believed to reduce these gaps. Government key informants referenced historical challenges regarding disaggregated data not being available citing that it was: “not a priority, not trendy”; programs not being designed with GBA/+ elements in effect; programs not having information available on who is using programs; and perhaps one of the most important concerns, the costs of constructing and maintaining systems to collect this information. Relative to this last point, government key informants also spoke to the need for data integrity to ensure confidence during analysis, and the crafting of appropriate recommendations. While some of the earlier GBA/+ concerns have diminished with the support of the current political leadership, others continue to plague the evaluation function particularly when elements of sustainability are introduced into the accompanying narrative. There is a fear factor related to significant investments in initiatives that may lead to a culture change when GBA+’s permanence is questionable. One gendered expertise interviewee identified “a fear factor” of subsequent findings may reveal. When regarding initiative beneficiaries as heterogeneous, different needs will emerge that the initiative may not be prepared to address with existing allocated resources.

Dishearteningly given that disaggregated data is a cornerstone of GBA+, only 38 percent of respondents to the HoE survey noted their organizations systematically collecting disaggregated data on multiple identity axes to support evaluation efforts. Interestingly, only 58 percent of respondents noted this as being problematic. This finding contrasts sharply with

findings reported from the annual GBA+ implementation survey's most recent findings. In these accounts, it was reported that in 2017, 83 percent of respondents noted that their organizations collected disaggregated data which was up from 77 percent in 2016 (Status of Women Canada, 2017) These discrepancies bear further investigation and lead to the requirement to clarify at what level of disaggregated data is being collected.

The missing accountability stick in Gender-based Analysis/Gender-based Analysis Plus (GBA/+) narratives

Policy studies repeatedly reference the need for some combination of carrots, sticks, and sermons for successful implementation (Bemelmans-Videc, Rist & Vedung, 1998). The “stick” mechanism is believed to be particularly effective because it introduces an accountability dimension into policy development and implementation. Unfortunately, GBA/+ has been missing the accountability stick. The argument could be made that the 2 Auditor General reports represent an accountability mechanism, but perhaps only for the departments involved in these audits of GBA. Identified shortcomings in the accountability narrative, however, stem as far back as 2005 with the Expert Panel on Accountability Mechanisms for Gender Equality report. The report had identified no existing obligations for departmental reporting on GBA and made a series of recommendations regarding accountability measures surrounding GBA that included (The Expert Panel on Accountability Mechanisms for Gender-Equality, 2005, p.46)

- Articulate a clear set of goals and anticipated outcomes and link them to major gaps in achieving substantive equality for women;
- Develop key indicators and benchmarks;
- Establish formal monitoring and reporting mechanisms;
- Increase financial resources allocated to SWC and other relevant departments; and

- Strengthen horizontal coordination of government actions including GBA activities.

The implementation of these recommendations was a resounding failure after the immediate tabling of this report. Shortly thereafter in 2007, SWC experienced a massive cut to their programming delivery arm that contributed to these implementation failures. The 2015 GBA OAG report tabled a follow-up recommendation regarding the periodic assessment and reporting on progress (Office of the Auditor General, 2015, s.1.62) of GBA implementation across departments. This subsequent recommendation appears in the GBA+ Action Plan (2016-2020) jointly published by SWC, Privy Council Office (PCO), and TBS. In 2016, the first GBA+ survey was administered to Deputy Ministers (DM) exploring barriers to implementation, improve understanding of the implementation and provide a baseline to measure future progress. It is comprised of 20 open-ended and dichotomous questions delivered to 30 departments and agencies (Status of Women Canada, 2016). Shortcomings of the first annual monitoring effort include not specifically identify the evaluation function (although research was), highlighting a missed opportunity in recognizing federal evaluators as potential producers of GBA/+ information.

The first annual monitoring survey administered at the DM level also highlighted opportunities to improve processes and implement mechanisms tracking progress within individual departments (Status of Women Canada, 2016, p.3). In the 2016 FEWO GBA+ report to Parliament, a recommendation was made that each department has in place by June 2017 a mechanism to track its implementation of GBA at the individual organizational level. Government responses agreed in theory to this recommendation, however, recognition of this across federal government organizations varies at the time of this writing. Just as federal organizations are required to respond to the annual GBA+ implementation survey, so too are

organizations required to report on their evaluation functions. To ensure a risk-based responsive approach to reporting on results across government, the Management Accountability Framework (MAF) was implemented in 2003 (Treasury Board Secretariat, 2009). Originally, *Evaluation Quality and Use* was designated as *Areas of Management* (AoM) 6, although it was removed following the release of the 2009 *Policy on Evaluation* (Treasury Board Secretariat, 2009) which had substituted a requirement for federal government evaluation functions to undertake a neutral assessment of their evaluation function once every 5 years (Treasury Board Secretariat, 2009). This neutral assessment function was continued as a policy requirement under the 2016 *Policy on Results* (Treasury Board Secretariat, 2016) however GBA+ is not included in MAF exercises. To respond to monitoring, and monitoring by the evaluation function, one of the case study organizations identified the potential for and included GBA+ to be incorporated into their neutral assessment of their evaluation function.

The *Policy on Results* (Treasury Board Secretariat, 2016) identifies the roles of HoE concerning recommendations and the subsequent development of Management Response Action Plans (MRAPs). While it may be early considering the state of GBA+ implementation in evaluation, the possibility exists to monitor the implementation of related recommendations in this manner. HoE are responsible for reporting to their organizations' Performance Measurement and Evaluation Committees (PMEC) on the state of implementation of evaluation recommendations annually (Treasury Board Secretariat, 2016, S.4.4.4.) which could easily accommodate embedding of GBA/+ in evaluation recommendations, MRAPs, and the extent of implementation by program areas. This is exactly what one case study organization is intending to do. Interestingly, no key informants raised monitoring of GBA/+ within federal organizations during interviews.

One department repeatedly flagged as a success story in the GBA/+ accountability narrative was the Department of Citizenship and Immigration (CIC, now Immigration, Refugees and Citizenship Canada [IRCC]). This federal organization has a legally based annual GBA/+ reporting requirement to Parliament (The Expert Panel on Accountability Mechanisms for Gender Equality, 2005; Office of the Auditor General, 2009; Standing Committee on the Status of Women, 2016). This legal requirement flows from the Immigration and Refugee Protection Act. This parliamentary reporting on GBA/+ “means that the Department absolutely must do an annual report and this automatically becomes part of a deputy minister’s way of evaluating performance”. This quotation also raises the question of tying performance to results of senior decision-makers. Several case study organizations have introduced dimensions of GBA+ into executive performance agreements (n=3) to which financial incentives are tied. Key informant interviewees highlighted that embedding GBA+ formally has manifest in both the evaluation and performance measurement functions regarding GBA+ and resources allocated accordingly to support its integration.

Defining necessary, adequate, and sufficient resources

The question of necessary, adequate, and sufficient resourcing has been continually raised although what is meant by resources requires a more nuanced discussion. Findings from the AGE evaluation presented in an earlier thesis chapter highlighted that achieving success required increased dedicated staff, tools, and training materials, a clearer roadmap in addition to financial resources to drive Canada’s GBA/+ efforts forward (Status of Women, 2005). Evaluation key informants identified that during this current era competing priorities for SWC expertise were problematic. Despite SWC’s staff possession of the necessary expertise to assist individual organizations, access to this expertise was difficult to obtain given increased demands. These

increased demands were perceived to be attributable to newly introduced GBA+ policy requirements. The latest annual GBA+ implementation survey highlights note (Status of Women Canada, 2018) an increase in staff resources within individual organizations with more than 5 full-time equivalents (FTEs) dedicated to GBA+. This number of gendered expertise FTEs represents an increase from 10 percent in 2016 to 43 percent in 2017. No organizations have 0 FTEs dedicated to GBA+ in 2017. Comparably, central agencies also monitor the state of the evaluation function. Data from the Capacity Assessment Survey (CAS) notes a steady decline in evaluation FTE's from 2013/14 to 2016-17 (Open Government, 2018). Unfortunately, the CAS does not report on evaluation resources dedicated to GBA+. CAS open access data also notes a significant decline in professional services funding available to the evaluation function, dropping from \$12.4 million to \$8.2 million per year over the same period. This is important as when various organizations' evaluation functions lack resources in-house, contracts are let under professional services to build internal capacities, or to undertake some or all of the evaluation work itself.

Individual federal organizations highlighted early on the requirement to tailor SWC developed training materials for use within their unique organizations. While SWC's online GBA+ tool is useful and has been used by many departments as the basis for their training, the lack of evaluation-specific content in these materials is problematic. Building specific tools for the evaluation function has also resulted in a small resource burden as individual departments have contracted for services to help build related capacities. When queried as part of the survey administered to HoE as to whether their evaluation function had the necessary financial resources to integrate GBA+ in their evaluation efforts, only 62 percent of respondents positively identified having access to financial resources to do so. What these financial resources might

encompass was unpacked by key informants who identified the following resources required: mandatory training beyond SWC's online GBA+ training, infrastructure requirements, organizationally specific evaluation tools, the acquisition of contextual expertise (for example, understanding how to integrate GBA/+ in a particular setting and with the various stakeholders involved), and perhaps most importantly, setting aside the time to properly research the front-end of evaluation planning and design to ensure that GBA/+ concepts and steps are correctly applied to evaluation processes and procedures. Only one key informant identified no financial resource burden. Specific reference was made to resources required to onboard program and senior management, the development of an organizational evaluation/GBA+ tool, and time to reflect after piloting efforts to improve their efforts moving forward.

Resources play an important role in GBA/+ implementation. The critical review returned evidence regarding the CIC/IRCC success story. The 2005 FEWO report cited the following on CIC's GBA resources: "the department's strategic framework identifies the ongoing requirement for ongoing reviews of dedicated resources to ensure ongoing abilities to meet legislated commitments" (The Expert Panel on Accountability Mechanisms for Gender Equality, 2005, p.33). Such observations accord a somber reflection upon what is required to successfully integrate GBA/+ and produce a quality product. The Expert Panel further extrapolates on resources requirements noting that "any department required to undertake gender-based analysis requires the additional monetary and human resources to make the process work. This goes with the precept that if we expect a government body to accept responsibility for an initiative, we must ensure that it has the capacity to fulfil the initiative" (The Expert Panel on Accountability Mechanisms for Gender Equality, 2005, p.29). The question arises naturally as to where these resources might come from? In Budget 2016, SWC was provided \$23.1 million over 5 years to

support GBA+. A portion of this funding was to flow to federal organizations to build GBA+ capacity but no funding was earmarked for the evaluation specifically function (Standing Committee on the Status of Women, 2016, p.10). It might be reasonable to divert some of this funding, particularly as noted by SWC itself, that the evaluation function is lagging behind other functional communities (Status of Women Canada, 2018). This aligns with Recommendation 10 made in the 2016 FEWO report regarding the request for additional resource allocations, including personnel and finances, be made by the federal government to support the development and provision of training materials (Standing Committee on the Status of Women, 2018, p.22).

Sustaining momentum

Highly intertwined with GBA/+accountability and necessary resources is the issue of sustainability. With the election of Justin Trudeau, what has occurred is essentially the rise of the third wave of GBA/+ in Canada. This third-wave GBA+ includes several new advancements that address in part, some of the shortcomings previously identified. With the rise and retreat of each new wave, some efforts and headway remain. As of yet, the rise of third-wave GBA+ has yet to demonstrate the value of GBA+ in evaluation efforts to consistently provide information to support decision-making. The various kinds of use of GBA/+ information and use are addressed in the following thesis chapter. Within the evaluative context, investments have been made but very little return on investment has yet been evidenced. The majority of federal government evaluation functions are just beginning to build capacity in this area, and useable data for decision-makers remains unknown. It is particularly noteworthy to acknowledge this concern given an upcoming federal election and possible change in political leadership. While not a mitigation strategy per se, with the rise and fall of each subsequent wave of GBA/+, an increased number of individuals, and functions within the federal government have been identified as

capacity-building targets. When the current wave recedes and gender equality is no longer a priority, the knowledge generated remains. Individuals with their gendered expertise will transition from their current positions into other functional areas across the government. The knowledge is then not lost precisely but rather, is integrated into other sectors and domains. Several key informants across the domains identified, and case study participants spoke to the transitional nature of careers in the federal public service.

Federal government interviewees were optimistic regarding the sustainability of efforts supporting the integration of GBA+ into the federal government evaluation function. References were made to the concrete articulation in the *Policy on Results* (Treasury Board Secretariat, 2016) and accompanying *Interim Guidance on Results* (Treasury Board Secretariat, 2017), and the integration into other policy junctures as enabling factors. Additionally, the accompanying value that could benefit the evaluation function was explicitly explored which includes the opportunity “to think really deeply about what different population groups and identity factors may be impacted by an intervention”. Being able to reflect on an initiative's reach to various stakeholders was believed to result in more gender and diversity information being provided, and thus increasing the quality of evaluations undertaken. One caveat was issued; that the integration of GBA+ in evaluation be done in a non-tokenistic manner, and thoughtfully. Interviewees also spoke to the engagement of key influencers in the federal government bureaucracy that served as enablers. The DOF was noted as being a significant stakeholder given their previous derelictions noted in tabled OAG reports; their current visibility and leadership in supporting the implementation of gender responsive budgeting signalled to stakeholders' advancements in GBA/+. For these reasons, the perception was held almost universally amongst federal

government interviewees that GBA/+, and its integration in the federal government evaluation function was more than just a trend.

At the risk of being rote

Given the mass-scale implementation of GBA+ across the federal government, a danger exists that such an exercise can be more symbolic than meaningful. This is particularly relevant for the evaluation function given the high demand for GBA/+ across various policy cycle junctures, the limited accessibility of GFPs and GC's expertise, and the limited expertise within the evaluation function. Critical review evidence directly referenced the dangers of symbolic application. The 2005 Expert Panel report noted that "public reporting is often good but it can sometimes become simply mechanistic" (The Expert Panel on Accountability Mechanisms for Gender Equality, 2005, p.24). Independent processes exist for both the GBA+ and evaluation functions; each following a specific order of related tasks with associated responsibilities and actions undertaken at each phase. Other potential downsides of a mechanistic application GBA+ may result in negative unintended consequences for various subsets of program beneficiaries. Case study participants universally noted the importance of engaging in the messiness and complexity that integrating GBA+ in evaluation entails. Embracing this messiness and avoiding mechanistic applications can be summarized with the following quotation from a key informant "but when we are talking about people, if you don't take this into consideration fully, and by fully I mean really incorporating, embodying their voices into the evaluation process itself, then there is nothing". (personal correspondence, 2018).

Avoiding the simplistic, rote, and mechanistic integration of GBA+ in evaluation is not easily addressed. Some evaluation functions participating in case studies (n=2) flagged that they had been mandated by their ADMs to include GBA+ in each planned evaluation. Other

participating case-study organizations have included GBA+ in evaluation efforts by highlighting policy requirements, and others still identified that they had integrated GBA+ in evaluation efforts because it made logical sense in serving all Canadians (n=1). Key informant interviewees spoke of the value of integrating GBA+ in evaluation and framed efforts as “solid”, “doing it right”, and “doing the right thing”. Oftentimes these qualities used to describe GBA/+ efforts are made without knowledge of the intersectional theories guiding GBA+. An unfamiliarity with theoretical dimensions of GBA+ (emphasis on the “+”) inculcates an emphasis on “intersectional” which starkly contrasts with GBA’s emphasis on gender.

As a result of this emphasis and critical to intersectional theory, is the role played by institutions and systems in perpetuating unequal access and participation. Both GBA+ and evaluation are uniquely positioned to challenge status quo assumptions regarding how people operate within systems, and how these systems even unconsciously perpetuate discriminations. However, positioning these functions in challenging the status quo, does not necessarily lead to change. This includes addressing existing social norms and values which are difficult to measure and are often lost in the nuances of public reporting. Reviews of evaluation reports from the case study line of evidence note that rarely do evaluation recommendations emphasize the role of federal systems in perpetuating discrimination. Several individual key informants representing the federal government perspective identified the value of having an in-house critical theory or colonial studies knowledge represented on evaluation teams as it was believed that these critical analysis skills would address institutional systems' roles and responsibilities.

References were made by a few case study participants to the neutrality/objectivity quality construct supposedly embodied by the evaluation function to demonstrate that these constructs were/are artificially constructed and that science has been used to illustrate very real differences

that exist. Presumably, the avoidance of mechanistic applications necessarily involves some demonstration of use which is discussed in greater detail in the following chapter of this thesis.

Quality of Gender-based Analysis Plus (GBA+)

The question of GBA/+ quality peripherally factored into both OAG and FEWO examinations of GBA/+. In the 2015 OAG GBA report, the emergence of quality criteria emerged. Assessments of quality were tri-dimensional and included: a review of relevant data sources, perspectives of affected groups were obtained, and the examination of all GBA considerations (Office of the Auditor General, 2015). After identifying these dimensions of quality, the report identified that the GBAs they had examined were incomplete. This conclusion was predicated on the lack of recommendations or options presented to address identified inequality gaps. Subsequent quality criteria discussions were resumed in FEWO 2016 report: the quality of GBA was compromised when it was applied unevenly and without a consistent framework (Standing Committee on the Status of Women, 2016). Despite the objective of providing information to inform decision-makers, having confidence that data is valid and reliable would undoubtedly factor into any decision-making process. GBA+ quality criteria have yet to be established for either GBA/+, and for the integration of GBA/+ in evaluation despite the federal government outlining its expectations for evaluation quality criteria. Evaluation quality criteria are outlined in *Appendix C: Standard on Evaluation* of the accompanying *Directive for the Policy on Results* (Treasury Board Secretariat, 2016). Appendix C outlines requirements for the evaluation function to address while planning, undertaking, and reporting on evaluations. No such comparable policy document accompanies GBA+. This is paradoxical given original SWC GBA training materials (Status of Women Canada, 1998). In these materials, the final step in completing a GBA included evaluating the quality of the work.

Additionally, these early training materials explicitly reference the respective roles of both the performance measurement and evaluation functions in the GBA/+ quality narrative (Status of Women Canada, 1998). In 2001, SWC referenced this relationship citing “[p]ressures for greater accountability to the public on the outcomes and the impacts of government programs on the lives of Canadians create the need to evaluate the effectiveness and efficiency of GBA and of its relevant processes in achieving actual gender equality results. There is also the need to evaluate the validity and the quality of GBA as a sound policy tool” (Status of Women, 2001, p.7). Over the years, this last step has been lost at the GBA/+ output level and for GBA/+ as a policy instrument. Interviewees spoke articulately on the relationship between quality and GBA/+. They referenced allocating sufficient resources (including capacity, time, expertise) to undertake quality GBA/+, the need to alert senior decision-makers to these resource requirements to undertake quality GBA/+, and earning the confidence of senior decision-makers to support making investments that would allow for undertaking quality GBA/+. In a few instances, the fear of having GBA/+ results made public was also referenced by federal government key informant interviewees “I’m afraid of is that it’s going to point out the fact that we don’t have the time to do very good GBA in general because of time, lack of time”. Undoubtedly, a quality GBA+ is one element of accountability that requires further investigation.

Participating case study organizations echoed these sentiments and spoke in greater depth on aligning GBA/+ results for various policy cycle inputs. Shared responsibilities between performance measurement and evaluation function for localized disaggregated data and its current unavailability were noted across all lines of evidence. The robustness of a GBA+ can be questionable if data isn’t available. Evidence also highlighted the need to determine what GBA+ data, and in what format this data exists, and what data should be shared with senior decision-

makers and other federal government functions. GFPs interviewed as part of case study samples also articulated the requirement for additional lead time to assemble and present data that could have potentially been generated by the performance measurement and evaluation functions.

Universally, challenges were experienced in producing a quality GBA+ output within what was articulated as “tight timeframes”. In one instance, an interviewee noted a 2 week time timeframe to provide several program areas with a GBA+ annex for a newly introduced policy requirement as part of the 2017 budget exercise. These short timeframes do not induce confidence in the production of a quality GBA+ product. While this example was provided by gendered expertise, the same concept can be applied to evaluations. Interesting dimensions that contribute to GBA+ quality were raised by key informants including:

- Knowing what GBA+ questions to ask, and when to ask them;
- Evaluators not knowing how to undertake intersectional analysis; and
- Being aware that senior decision-makers don’t have ready access to information generated from the consultation process. Consequently, some findings may not be included in the GBA+ outputs being reviewed.

Responsibility for producing a quality GBA/+ was raised in the critical review line of evidence. In the FEWO 2016 report on the state of GBA+ implementation, it was suggested that senior management “should review and challenge the quality of GBA+ with the goal of strengthening this portion of employees’ analysis. (Standing Committee on the Status of Women, 2016, p.23). Elevating the responsibility for assessing GBA+ quality to this level would ensure the prioritization of this message was heard across the organization but as previously discussed, historical records reflect resistance to this suggestion.

When interventions are grounded in Gender-based Analysis Plus (GBA+) elements

The evolution of GBA into GBA+ has to an extent left a historical legacy that unintentionally focuses on sex and gender as key analytical variables. Until January 2019, SWC's GBA+ public-facing materials positioned sex and gender as the entry point into intersectional analysis. Other identity variables were to follow. Just recently, amendments have been made that allow for the positioning of other identity variables as primary, or entry point analytical variables in analysis efforts. Reasons for the change in approach were not publicly announced but the caveat provided on SWC's website attributed this shift to alignment with departmental mandates. Some federal government programs target specific GBA/+ dimensions such as women, disability, Indigenous or Aboriginal, seniors, etc. Some of these programs are housed in federal government organizations dedicated to serving these populations (such as Status of Women [SWC] or now Women and Gender Equality [WAGE] and Indigenous and Northern Affairs [INAC] now Indigenous Services Canada [ISC] and Crown Indigenous Relations and Northern Affairs Canada [CIRNA]). Other federal organizations support programs that have these programs embedded within their portfolios.

Respondents to the HoE survey were specifically asked whether any of the evaluations that they had undertaken were embedded in specific elements of diversity. Fifty-seven percent of respondents noted that some of their evaluations were focused on elements of gender or diversity. Additionally, when queried on whether they had handled GBA+ differently, a minority (n=3) reported that they had indeed handled it differently. Differential handling was also noted in the critical review, key informant interviewees, and participating case study organizations' lines of evidence. When those programs were grounded in EDI principles, evaluation efforts did not apply GBA/+ as originally conceptualized. The rationale highlighted that because initiatives were grounded in EDI, GBA/+ was not necessary.

When queried as to whether these evaluations had captured individual level GBA/+ data as required, the response was negative. This thinking poses an interesting conundrum and raises the question of whether it is appropriate to collect data at this level. In the fall of 2018, the former federal government department INAC had identified on their website that they were currently reflecting on what constituted *culturally appropriate* GBA+. Several participating case study organizations (n=3) had highlighted that it was problematic to collect GBA+ data from Aboriginal communities at the individual level. Cultural norms emphasizing Aboriginal beliefs regarding the holistic community were provided as a rationale. These organizations also noted that their evaluation processes related to data collection may also be culturally inappropriate and insensitive to the lasting legacy of colonialism (i.e. brief visits by the evaluator for one specific evaluation in remote communities). Future research in this area in light of recommendations flowing from the Truth and Reconciliation Commission (TRC), and the First Nations' OCAP (ownership, control, access, and possession) principles concerning knowledge production relative to GBA+ would provide rich fodder. Potential findings may also apply to other federal government targeted populations.

Chapter 6 - Dear Betty: My cake is out of the oven. What do I need to decorate?

This thesis chapter presents findings related to the capacity to *produce* GBA+ specific to the evaluation function. The discussion begins with what constitutes a completed GBA+ and continues by exploring how GBA+ may be integrated in evaluation processes and products inclusive of evaluation core issues, evaluation frameworks, lines of evidence, and phases.

Capacity to produce Gender-based Analysis Plus (GBA+)

Evidence from the critical review identified that a GBA/+ is considered completed if an organization has (Standing Committee on the Status of Women, 2016, p.6):

- Reviewed data sources such as research papers, quantitative data such as statistics, and other gender-related information relevant to the initiative;
- Considered the perspectives of stakeholders, including affected groups of women and men;
- Examined the gender considerations raised in data sources and by stakeholders, to determine whether any are relevant to the initiative and could have different impacts on specific groups of women and men; and
- Where relevant gender considerations exist, proposed options or risk-mitigation measures to address any inequalities in the outcomes of the initiative on those gender groups, or justified why the gender considerations do not apply to the initiative.”

The federal government’s evaluation function is being asked to do more than these four criteria. Solid arguments can be made that the evaluation function flexes its analytical capabilities well beyond the completion criteria identified above which include creating data sources from multiple and varied stakeholders to determine differential outcomes. Evaluation functions are not only reviewing data sources but also producing both qualitative and quantitative information by

various identity factors relevant to the initiative. The evaluation function is tasked with not only confirming who affected groups might be but oftentimes is asked to identify alternatives.

Arguments were raised across key informants that established evaluation functions would undertake these activities normally as part of a normal evaluation, particularly if methodological affiliations were aligned with participatory, collaborative, empowerment, transformative, gender, or feminist orientations. Although these instances exist in the federal government context, they are not frequently employed. An example provided by one key informant referenced Global Affairs Canada's (GAC) distinct feminist monitoring and evaluation projects piloted in the international development domain and the alignment with the feminist development agenda.

Conceptualizing the capacity to produce GBA+ in organizational evaluation functions commences with a perception of reasonableness. Respondents to the HoE survey reported at 91 percent that within their organizations, that it was reasonable to integrate GBA+ given the current state of their evaluation function. When reporting on actual integration, however, the difference between perceptions and actual implementation decreases to just 36 percent when queried on consistent GBA+ integration. This reporting is significantly different, and much lower, from a similar finding reported in SWC's annual GBA+ implementation survey. This difference between the two implementation accountings is approximately 45 percent. Minor differences may be accounted for when considering target audiences, and parsing reporting responsibilities to appropriate respondents however, this is a major difference. It is suggested that HoE are best positioned to respond knowledgeably to their evaluation functions' activities and that the findings represented here are accurate. Federal government key informants provided information from the SWC 2016-17 annual GBA+ implementation survey findings relative to the evaluation function. The state of GBA+ integration was summarized as: "[O]verall, GBA+ was

not strongly integrated. GBA+ in evaluations made public in 2016-17, almost two-thirds of organizations reported that few (0-20%) of publicly available evaluations identified the impacts of programs or policies by gender and other identity factors” (SWC, personal correspondence, 2018). Inferences were made based on the addition of GBA+ to the *Policy on Results* (Treasury Board Secretariat), that evaluations would immediately begin integrating GBA+. This did not consider that federal government evaluations often require months, or even years of planning, analysis, and reporting. This suggests a misalignment between any expectations and understandings of how the federal government evaluation function operates. Respondents to the HoE survey also were asked whether they believed they had the necessary financial resources to support integrating GBA+ in their evaluation function. Only 62 percent of respondents affirmatively noted that this was the case. Differences in findings between perceptions surrounding reasonableness, consistent implementation, and resources highlight gaps in implementation.

Respondents to the HoE survey were asked to place their evaluation function’s capacity to integrate GBA+ in evaluation along an intersectional analysis continuum that included junctures identified as “low”, “developing”, and “established”. Most respondents (n=13) highlighted their evaluation function as “developing” with “low” scoring second (n=5). No respondents identified their evaluation function’s capacity to integrate GBA+ as established. Further, respondents were also asked to identify on an intersectional research continuum where their evaluation function currently resided, and where they believed the evaluation function might be in 2 years. These five intersectional analysis continuum points and broad definitions were included as response options:

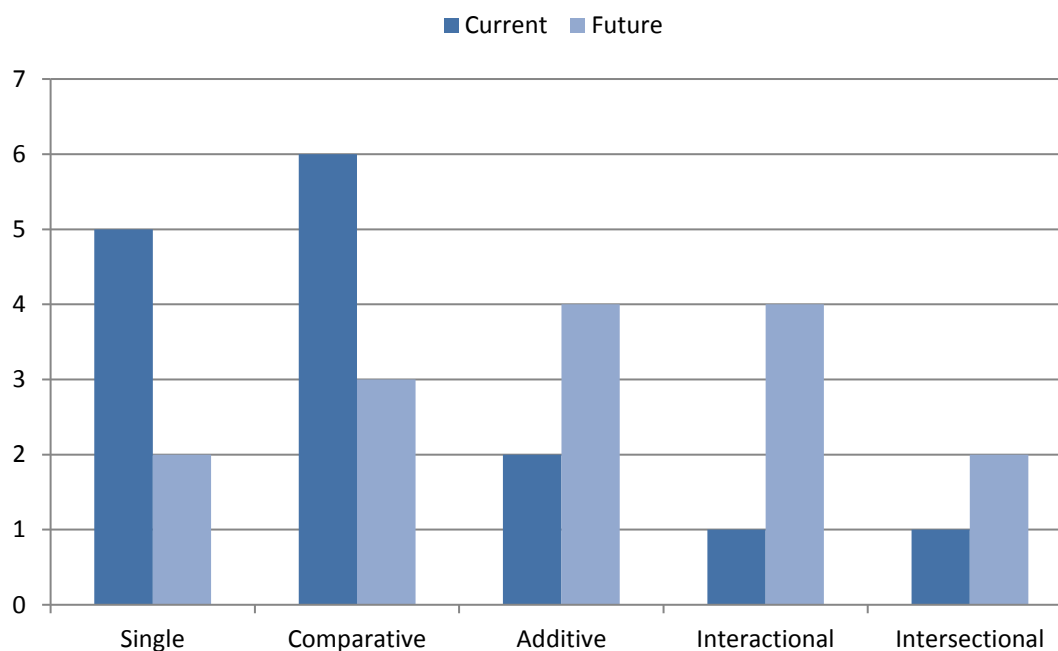
- Single (one identity variable);

- Comparative (one identity variable, limited categories);
- Additive (more than one identity variable, limited categories);
- Interactional (more than one identity variable, multiple categories and inclusive of context); and
- Intersectional (multiple identity variable, multiple categories, and inclusive of context).

Only the last option presented reflects the desired state of GBA+ as truly intersectional analysis.

Figure 7 presented below depicts the current and future projected state of intersectional analysis by the evaluation function.

Figure 7 Current and future projections of evaluation function intersectional analysis



Most organizations' self-reported current capacities reflect single and comparative intersectional research junctures which are aligned with low and developing capacities. This reflects the historical legacy left by GBA in its transition to GBA+ and echoes the sentiments expressed by key informant government interviewees with functions possessing the capacity to report

predominantly on sex and/or gender differences only. Very few respondents flagged their evaluation functions as having more established current capacities to integrate GBA+ in evaluation. Unsurprisingly, when regarding the future state of capacity, there is a general shift to the additive and interactional intersectional research junctures. Very few respondents overall however identified that their evaluation functions would have the necessary intersectional analytical capacity to undertake GBA+ in 2 years. This suggests several reflections on the federal government's evaluation function. Immediately, an increased timeframe may be required to actualize intersectional analysis capacities as well as creating the enabling environment supports and collaborate with other functional communities involved.

Consideration of the demand for GBA+ information was also explored in this research. HoE survey respondents were queried on both programs' and senior decision-makers' demand for gender and diversity in their decision-making processes. HoE survey respondents reported at 81 percent that their senior decision-makers considered gender and diversity in their decision-making processes. This reported rate varies from program decision-makers' considerations of gender and diversity in their decision-making processes noted at 67 percent. This variance of approximately 20 percent speaks to the influence of political will in considering GBA+ given that senior decision-makers want to ensure that those policies, programs, processes, and interventions under their jurisdiction are meeting federal government priorities. Program management may not experience political will in the same way as senior decision-makers. HoE survey respondents were also asked to reflect on whether the program and senior decision-makers have specifically asked for the inclusion of gender or diversity considerations in evaluation efforts. Findings in these two areas dropped significantly; only 45 percent of senior decision-makers were recalled as having specifically requested gender or diversity be integrated

in evaluation. Similarly, a comparable decrease was also recalled for program managers, moving to 38 percent having specifically requested gender or diversity being integrated in evaluation efforts. Determining evaluative direction at the onset of evaluation planning efforts is frequently a negotiated process between the various stakeholders involved (Bourgeois & Whynot, 2018a). Should either program or senior management of an organization request that an evaluation include GBA+ elements it would be considered by the evaluation team, particularly if the request came from senior management. There are many ways by which to include GBA+ in federal government evaluations that include integration of GBA+ in the core issues, in the evaluation framework, and within various lines of evidence. Findings from each of these three elements are discussed in greater detail below.

Both government key informants and participating case study organizations noted that moving from considering program beneficiaries as homogeneous to heterogeneous was a significant revelation. In recognizing that program beneficiaries were heterogeneous, there was an accompanying identification that program beneficiaries' needs would also be different. Initiatives were scaled based on meeting the needs of *most* Canadians, rather than focusing on meeting the needs of *all* Canadians. As such, barriers and challenges would be experienced by some Canadians at both access, and participation stages of these interventions. Programs could potentially engage all Canadians by developing tools and processes that addressed barriers to their access and participation which would lead to improved access and participation in government initiatives.

In both government key informant interviews and participating case study organizations, the transition from reviewing program beneficiaries as homogeneous to heterogeneous was novel. This recognition was universally attributed to GBA+ efforts although those case study

organizations providing direct services to individuals had a much easier time identifying who program beneficiaries were. Participating case study organizations found it easier to identify barriers when applied to a single identity factor such as gender, but when including multiple categorizations such as Indigenous, Muslim, elderly, etc. complications arose. Additional tensions were experienced when attempting to formulate related recommendations that were attributed to the specificity of program beneficiaries against the general nature of the program.

Integrating Gender-based Analysis Plus (GBA+) in evaluation core issues

Decisions made on *where* to integrate GBA+, and at *what level* have direct implications for *how* GBA+ is integrated in evaluation efforts. Evaluation frameworks are used as a tool to structure federal government evaluation efforts. Key evaluation questions form the foundation of each evaluation framework. Key evaluation questions correspond to evaluation core issues. The legacy left by previous federal government evaluation policies has resulted in core issues to be addressed by federal government evaluation efforts. These core issues are related to assessments of the relevance and performance of government programs. Relevance core issues include the following (Treasury Board Secretariat, 2009):

- The continued need for the program (addressing a demonstrable need and responsive to the needs of Canadians);
- Alignment with government priorities (linkages between program objectives and federal government priorities as well as departmental strategic outcomes); and,
- Alignment of federal roles and responsibilities (roles and responsibilities for the federal government in delivering the program).

Participating case study organizations highlighted the natural alignment of integrating GBA+ in evaluation with the first core issue. This is particularly evident when referring to the

“needs of Canadians”. Universally, case study participating organizations noted that GBA+ should be lauded for its persistence in pushing for the acknowledgment that Canadians are diverse, and accompanying this diversity are different needs that programs need to respond to. This varies from historical legacies in viewing Canadians as homogeneous groups with little accord for diversity, never mind diversity within diverse groups. Organizations such as the Public Health Agency of Canada (PHAC) and Health Canada (HC) who share an evaluation function, have been mandated by their Assistant Deputy Minister (ADM) to include three GBA+ related questions in each of their evaluations (Public Health Agency of Canada, 2019) for the foreseeable future. Proposed questions include addressing the first core issue in evaluation. Cited potential benefits of integrating GBA+ included the potential for change amongst diverse Canadians. One government key informant noted a fear expressed by colleagues related to asking questions at this level. This was because presumably, returned information would have significant financial implications for programs having to implement changes within their programs to meet the needs of diverse groups of individuals. As may be recalled, few programs were designed with the diverse needs of all Canadians in mind. The critical review reported no information on this potential resourcing concern in implementing changes resulting from the use of GBA+ information.

On the second core issue, assessing the linkage between program objectives and federal government priorities as well as departmental strategic outcomes highlighted several challenges. A range of opinions was expressed by key informants and case study participating organizations reflecting on past, current, and future government priorities. The first of these concerns flagged that while GBA/+ is widely mandated across the policy cycle, not every program moving forward had included GBA/ in associated TB Subs and MCs. Gendered expertise key informants

indicated that in most instances, policy cycle junctures requiring GBA/+ input were completed under time constraints. Subsequently, it was perceived that the quality of these GBA+ components was often compromised as a result. Noted compromises included using dated (not recent) GBA+ information, missing personal identity elements, or those that are incorrect. One broad concern of GBA+ lies with author responsibility as these individuals hold the power to define inclusion/exclusion of stakeholders and their perspectives.

Historically, policy cycle junctures requiring GBA/+ input were often summarized as “not applicable”, or gender neutral. Critical review documentation evidence (Standing Committee on the Status of Women, 2016) spoke to these concerns suggesting to central agencies in their GBA+ challenge functions to return all such proposals. Central agencies were directed to refuse and return any such proposal to submitting organizations. While the status of this recommendation is currently unknown, the opinion was forwarded by several key informants that requesting programs to report on GBA/+ in evaluations was problematic. They countered that holding programs to account for GBA/+, when GBA/+ was not considered in program design and/or delivery was ill-conceived and contrary to principles of evaluation. It was argued that programs were never conceived to be responsive to the needs of diverse groups of individuals. GBA+ considerations would then never be reflected in program priorities, or in departmental strategic outcomes which form the basis of the second core issue. Evaluators from case study participating organizations managed this concern by recognizing program concerns as valid but drew on articulated requirements in the *Policy on Results* (Treasury Board Secretariat, 2016). They advised programs that evaluation efforts would examine GBA+ sensitively and would position evaluation recommendations that would assist programs in aligning GBA+ efforts to address gaps. Evaluators also brought to light the potential for programs to request

additional funding to respond to potential evaluation recommendations regarding programs' capacity to undertake GBA+. On the third core issue of aligning federal roles and responsibilities (assessment of the roles and responsibilities for the federal government in delivering the program), little discussion was offered on this point.

Core issues to be examined in evaluation can also be summarized as performance-related. These performance aspects include (Treasury Board Secretariat, 2009):

- Achievement of expected outcomes (assessment of progress towards expected outcomes with reference to performance targets and program reach, program design including the linkage and contribution of outputs to outcomes); and
- Demonstration of efficiency and economy (assessment of resource utilization in relation to the production of outputs and progress towards expected outcomes).

Significant attention has been paid to the performance core issue which can be explained in part by GBA+'s focus on impacts. Most lines of evidence highlight that federal government organizations are currently unprepared to fully integrate GBA/+ in the performance core issue. This is in part, attributable to GBA/+'s focus on impact; if individual evaluations are to report on impact, supporting information must be available, not only at this level, but at preceding results chain phases including inputs, activities/outputs, and subsequent outcome levels that include immediate (short), intermediate (medium), and further. As previously noted, the reach of most federal government evaluations extends predominantly to the intermediate outcome level given the 5year program renewal cycle. For this reason, the importance of any programs' assumption of responsibility for both GBA/+ and performance is increasingly emphasized.

Key informant interviewees spoke with varying degrees of familiarity on the various aspects of the performance core issue. Reach as a uniquely evaluative concept (Montague &

Porteous, 2013) emerged as a dominant theme particularly as it related to an intervention's design and delivery. Reach is evaluatively synonymous with GBA+'s recognition that Canadians are not a homogeneous entity and have different needs. Reach as a distinct layer in logic models, theories of change, or PIPs, has not yet been universally embedded across the federal government's evaluation function. Academic key informants also flagged that authorship in defining reach was potentially problematic as it is those individuals functioning within the power systems supported by federal governments and its institutions rather than the diverse Canadians who would be crafting initiative results chains. As a result, a privileging of those individuals' knowledge and perceptions about who might be affected by reach might occur which is similar to the privileging of GFPs' knowledge identified by Scala & Patterson (2017).

Evaluation framework levels and Gender-based Analysis Plus (GBA+)

Conceptually, questions were raised as to whether GBA+ should be a key evaluation question, an accompanying sub-question, or relegated to indicator status within evaluation frameworks. Evidence varied widely on this subject and was found to be largely dependent upon unique attributes of the organization, its evaluation function, and the evaluand's focus. Key informant and case study evidence suggested debates as to whether to embed GBA+ directly or indirectly in each of the possible levels within evaluation frameworks. Several interviewees including key informants and participating case study organizations identified the direct approach was associated with transparency in meeting federal government priorities. The direct approach was also perceived as having a degree of methodological control by those with more experience in integrating GBA+ in evaluation. Indirectly integrating GBA+ in evaluation frameworks at various levels was found to be associated with an era in which GBA+ was not a

federal government priority. Additionally, the tendency to indirectly approach integrating GBA+ was to “hide GBA+ in indicators” without being explicit that this was what was being done.

Case study participant organizations varied in the extent of integration, and level of integration of GBA+ in their respective evaluation frameworks. We begin this discussion with a closer examination of integrating GBA+ when evaluations are grounded in elements of diversity and inclusion as it would appear that the application of GBA+ in these instances is quite different. Earlier in this dissertation, it was identified that the Women’s Program at SWC did not apply GBA+ as conceptualized; meaning that the unit of analysis was not drilled down to the individual although intersectionality was introduced in various lines of evidence including surveys and key informant interviews. Key informants from the federal government and gendered expertise noted that when evaluations are grounded in elements of diversity, they should be approached similarly to how all federal government evaluations are approached (prior to the federal government’s articulation of GBA+ as a priority). Respondents to the HoE survey were asked two relevant questions that included:

- Of the evaluations that your organization's evaluation function has undertaken, do any of the interventions specifically target an element of diversity?; and
- Was GBA+ handled differently in these instances?

Responses to the first question were surprisingly reported at a much higher incidence rate than anticipated. Almost sixty percent of respondents referenced evaluations undertaken within their jurisdiction had embedded specific elements of diversity. A small number of respondents (n=3) noted that GBA+ was handled differently when evaluations target specific elements of diversity. The first finding presented suggests that even though the presence of diversity in program delivery is not an anomaly, dealing with it *well* is. Feminist evaluation academics (Bustelo; 2016;

Espinosa, Bustelo & Velasco; 2016) have begun fleshing out what is required for evaluations of gender mainstreaming initiatives. These authors identified that both process and outcome criteria are necessary but at this time, a competency framework identifying necessary components and related criteria does not exist.

The inclusion of one case study organization whose evaluation focus targeted specific elements of diversity was intentional. Meetings with evaluation personnel and senior managers during the development of the evaluation framework highlighted where it might be possible to integrate GBA+ in this evaluation. Case study interviewees identified that despite being grounded in elements of diversity, GBA+ was not integrated in evaluation efforts as originally conceptualized. For the program, EDI was the primary focus and at the evaluation framework level, the level of analysis highlighted community versus individuals. When queried about not collecting data at the individual level, it was explained that program delivery was generally administered through third parties community organizations, and agreements to facilitate individual-level reporting through these organizations were not established in light of streamlining reporting suggestions as directed by central agencies. This situation is like the WP evaluation at SWC, and many other grants and contributions (Gs&Cs) programs across the federal government.

Participating case study organizations who had implemented GBA+ in their evaluation frameworks (N=2) elected to do so at the indicator level. Integration was facilitated after including GBA+ as a distinct feature in exploratory evaluation scoping exercises. In most instances, scoping efforts result in the production of an evaluation methodology report. In turn, these methodology reports are presented at the organization's Performance Measurement and Evaluation Committees (PMEC) for refinement, and approval. After this approval, relatively

minor changes only may be made to the evaluation framework. An example of indirect indicator level GBA+ application can be illustrated by the case study organization electing not to include GBA+. This organization ultimately ended up including GBA+ indirectly as an indicator under one of the relevance core issues related to needs and information was gathered in the survey line of evidence. Other case study organizations focusing on building capacities included GBA+ at the indicator level after careful discussion, and elimination of its possible inclusion at key evaluation, and sub-question levels. This organization's overall approach to integrating GBA+ varied significantly from other organizations. They had elected to standardize GBA+ related questions into each potential line of evaluation inquiry (surveys, key informant interviews, and focus groups) in all future evaluations. Additionally, this evaluation function had also integrated a cross-organizational evaluation of EDI into their 5-year DEP. Difficulties have been encountered as several leading personnel have left the evaluation function for other functions/departments. This personnel departure resulted in the loss of knowledge and momentum generated in this area which is not atypical within the federal government evaluation function. These individuals do however take this expertise with them to their new positions.

Other case study organizations electing to participate in this research with more developed capacities shared different concerns. In these instances, GBA+ indicators were included in organizational/program Performance Information Profiles (PIPs). This information was culled on an annual basis; discussions were raised in case study workshops/information exchange sessions regarding how best to use and present culled GBA+ PIP information. Discussions highlighted that evaluator needs varied from those of decision-makers. Editing GBA+ information to a manageable level was perceived to be necessary as lengthy reports would not be useful to senior organizational leadership. One of the most significant challenges faced by participating case

study organizations was for evaluation functions who engaged with programs that had no GBA+ experience, either in the program, or evaluation context. In this instance, programs had never completed the GBA+ annex for either TB Subs, MCs, or budget proposals or if they had, they might have entered the infamous phrase “not applicable”. In these instances, or if delivering sector initiatives, programs would be unlikely to know exactly who program beneficiaries are, never mind their needs, tools, and supports required to increase equitable access and participation. In these instances, the organization’s evaluation function was left to produce some of this information during an evaluation rather than relying on data generated by the performance measurement function.

Lines of evidence and Gender-based Analysis Plus (GBA+)

Methods and sources outlined in evaluation frameworks are often referred to as lines of evidence. The *Policy on Results* and the accompanying *Interim Guidance on Results* (Treasury Board Secretariat, 2016; Treasury Board Secretariat, 2017) identify requirements for both qualitative and quantitative lines of evidence in evaluation. Lines of evidence typically presented in federal government evaluations include literature reviews, reviews of program documentation (including performance measurement data), environmental scans of undertakings in other jurisdictions, key informant interviews including subject matter experts, focus groups, and surveys (of targeted evaluation stakeholders such as program beneficiaries, funding applicants, advisory group members, etc.). Additionally, there has been some movement towards social media analysis such as Facebook, Twitter, and Instagram under the “big data” movement. While each evaluation effort is unique with different evaluation foci reflecting individual organization needs, some combination of the above lines of evidence can be found in every evaluation.

Several factors drove an initial focus on producing GBA+ quantitative data. This includes the previously discussed emphasis on impacts, as well as SWC's initial GBA/GBA+ capacity building efforts with socio-economic organizations providing economic programming. It has been only in the last couple of years that the emphasis has shifted from quantitative data to include qualitative information as well (Status of Women Canada, 2018). This may be attributable to the recognition of the GBA+ dimensions and related requirements to address components beyond impact. Very few organizations at this point can consistently produce GBA+ quantitative information based on performance measurement information; only 2 of the organizations currently have this capacity, having been integrating GBA+ in either their evaluations or program areas for more than a decade. To undertake truly intersectional analysis, access to performance measurement data is necessary. Other participating case study organizations are moving in this direction; during this capacity building timeframe, the survey line of evidence has also driven the production of quantitative data for evaluation purposes.

Findings from this research unanimously reference the need to reframe traditional evaluation approaches to successfully integrate GBA+ in evaluation. These adjustments include how both qualitative and quantitative information might contribute to each of the key evaluation questions. The potential exists to integrate GBA+ in each line of evidence used in an evaluation; however, some alignments appeared to be more natural than others. For example, literature reviews, reviews of program documentation, and environmental scans of practices in other jurisdictions were particularly well aligned with explorations of program beneficiaries and their needs, determining the existence of barriers, and impacts for various identity categories. In evaluations conducted by case study organizations, several innovations regarding program beneficiaries and their needs were piloted using program documentation. These innovations

included merging databases from several of the organization's programs to ensure that quantitative data on program beneficiaries was available, to some extent their needs, and their access to, and participation in programming. Other database-related innovations included the integration of databases across organizations. These innovations are unique because databases are uniquely structured to meet the needs of their programs. Consequently, the information contained in one program's database, even within the same organization, will vary from another program's database. To merge these databases, significant efforts were invested in cleaning or aligning key personal identifier variables such as names, gender, sex, etc. plus other relevant information to the evaluation.

Qualitative lines of evidence were found to be aligned well with the following GBA+ dimensions: exploring different needs of program beneficiaries; the nature of the barriers; and tools and supports required by diverse populations to facilitate their access and participation in various federal government interventions. Innovations in qualitative methods included over-sampling of underrepresented diverse populations for focus groups that attended to the nature of the barriers encountered in accessing and participating in government programming. One of the case study organization's recognition of the need for greater information regarding the nature of the barriers encountered led to a methodological shift from what was originally envisioned. Focus group composition shifted from a mixed-gender to women and Indigenous women-only focus group. In these focus groups, significant effort was invested in developing relevant prompts and probes relevant to the engagement of specific diverse populations. Findings resulting from the over-sampled focus group recognized that women faced similar financial and non-financial barriers as men but also included gender specific barriers including those related to childcare and maternity, as well as those barriers regarding entering a male-dominated workforce

including sexist hiring practices. Findings from these restructured focus groups have been emphasized as particularly significant by the organization's evaluation function, and publicly presented to the evaluation community.

Case study organizations also invested resources in exploring what efforts and results were available in other jurisdictions (beyond their unique organizations and internationally). Evaluators who had scoured the academic and grey literature found it helpful to have undertaken this exercise. They spoke of being "comfortable", and "knowing the program or sector" when this background research was undertaken to position their efforts, and to justify their line of inquiry when questioned on the relevance of integrating GBA/+ in evaluation. This evidence was supported by key informant interviews.

Surveys are another line of evidence employed in federal government evaluations. Case study organizations that employed surveys in their evaluations targeted both program beneficiaries as well committees used by the program. Both participating case studies and key informant interviewees noted mixed results in integrating GBA+ in surveys. Surveys were found to be helpful in that they could easily capture information from a large number of recipients with relatively little investment. Collecting GBA+ information from program beneficiaries requires voluntary disclosure of what might be considered vulnerable information. Disclosing GBA+ information without providing appropriate information regarding its use and without appropriate safeguards in place could potentially place those engaging with federal programs at risk by exposing perceived vulnerabilities.

During a study-wide organization case study discussion debrief, what emerged was the requirement to be upfront about why the data was being collected, and how it was going to be used. This included positioning GBA+ as a tool to address systemic barriers encountered by

diverse groups of individuals. In the two instances in which GBA+ was integrated into evaluation surveys, insufficient responses were received to allow for statistically significant analysis. In one instance, the evaluation clearly identified the intention in their requests for GBA+ information, and how any resulting information would be used but potential safeguards were used. In debriefing sessions, the role of organizational leadership in supporting GBA+ efforts was highlighted and believed to be important to include in messaging to survey respondents.

Case study key informants also spoke to the importance of organizational leadership. They simultaneously highlighted tensions between central agency requirements to streamline funding recipient reporting burdens and having sufficient GBA+ information for evaluation purposes. These discussions suggested that the decision ultimately needs to be made by organizational leadership and others who set organizational priorities. This theme emerged during the participating case study organization debrief undertaken in June 2018. Some leadership cadres are unaware of the potential of GBA+; they require information and related examples which are in short supply within the Canadian context although they are rapidly evolving. It was suggested by both case study key informants in three of the case studies, as well as by key informants that integrating GBA+ into performance management agreements for executives would be a significant indicator of the seriousness of the commitment.

Some participating case study organizations have experienced moderate success in collecting GBA+ information in different ways, and at different times. In these organizations, GBA+ information is collected from individuals at the funding application phase. This means that applicants provide GBA+ information before funding is awarded. Applicants are asked a series of GBA+/EDI questions regarding gender, visible minority status, Aboriginal status,

disability, and language alongside the option “Prefer not to respond” for each identity element question. Ultimately this information would be used to assist in scoping evaluation efforts, as well as a program documentation line of evidence during the evaluation. The timing of the collection of this information is an important consideration given that potential applicants may experience pressure to respond to these questions to receive funding. Additionally, neither participating case study organization has initiated efforts to validate this information at the time of evaluation which case study key informants identified as the major challenge for GBA+ in their context (self-reporting). Importantly, the collected data from one of these case studies was used as the analytical basis to highlight success rate differences between women and men awarded funding. These findings were used to alter how the program delivered its programming to decrease gender bias and was presented in high influence factor academic journals. Interestingly, these efforts were initiated by the program rather than the evaluation function. Key informants from these organizations also highlighted the capacity of the evaluation function to influence programs to better support EDI initiatives and pointed to concrete organizational examples.

Integrating Gender-based Analysis Plus (GBA+) in evaluation phases

Evidence from the critical review noted that GBA+ should be integrated into all evaluation phases including planning/design, conduct, analysis, and reporting phases. In short, incorporating GBA+ into evaluation processes is about ensuring that opportunities exist to hear from, consider and report back from individuals who may have been traditionally omitted from evaluation processes. A methodological checklist sensitive to new GBA+ requirements across the policy cycle was developed for the evaluation function to ensure that GBA+ was considered across

evaluation phases. The checklist was developed using findings across various lines of evidence from this research. This checklist can be found at Appendix G of this dissertation.

Chapter 7 - Dear Betty: I made the decorations and assembled the cake. Who is going to eat it?

This chapter begins by debunking assumptions underpinning the tailored ToC developed for this research before summarizing the highlights of GBA+ instrumental, conceptual, process, and symbolic GBA+ uses identified in this research.

Debunking assumptions

Descriptive statistics (counts, percentages, and frequencies) were presented in previous chapters of this dissertation. While helpful in identifying various attributes of a phenomenon, this level of analysis does not address relationships between variables. In this chapter, a more integrated analysis is undertaken to address this issue. Chi-square tests, as well as Phi and Cramer's V, were employed when expected cell counts were <5 . Statistical analysis was managed using IBM's SPSS software which is specifically designed for analyzing quantitative data. In summary, just under 30 variables across the tailored theory of change were found to be more statistically significant than others using a $>.05$ threshold to ensure a confidence interval of 95 percent.

Seven conceptual categories of variables were identified that included awareness, policy influence, enabling environment, tools and training, capacity, decision-making, and intersectionality discussed in previous thesis chapters. Each of these conceptual categories contained at a minimum 2 associated dimensions and one conceptual category had 8 associated conditions. The conceptual categories and associated dimensions are presented in Table 7.1 below. Each of these conceptual categories and associated dimensions was examined statistically for their relationship with the mid-range outcome "consistent integration of GBA/GBA+ in evaluation products/processes". Multiple reasons existed for selecting this mid-range outcome that included:

- Given the relative nascence of GBA/+ requirements for the federal government evaluation function, longer-term outcomes identified in the tailored ToC will require time to achieve given evaluation planning timelines (that are reflected in organizational 5-year Departmental Evaluation Plans [DEP]). Holding that these outcomes would have been realized within the research period would have been extremely ambitious. Consequently, the selection of a mid-range outcome from the tailored ToC highlighted what may have been realistic to achieve within parameters available. It would be worth exploring in future research the relationship between these variable and decision-making level outcomes but it was beyond the scope at this juncture in research efforts;
- The perception was held that the outcome identified above is a precursor outcome to informed decision-making by both program managers and senior decision-makers. GBA+ process, conceptual and instrumental uses in some form must be present to make related information stemming from evaluation efforts available to support decision-making;
- Currently, no quality criteria have been established to assist in determining what constitutes “good” integration of GBA/GBA+ in the federal government evaluation function. Projections from the evaluation discipline that suggest for evaluation to be good, it must be useful (Patton, 2008). As discussed in the previous chapter of this dissertation, various evaluation GBA/GBA+ uses are possible. Ultimately, the *Policy on Results* (TBS, 2016) Section 3.2.4 notes that “Parliamentarians and the public receive transparent, clear and useful information on the results that departments have achieved and the resources used to do (TBS, 2016). This emphasis on instrumental use discounts other uses that may contribute to quality criteria and standards for integrating GBA/GBA+ in evaluation. As a consequence, the outcome related to consistency was

used. Findings presented in this dissertation represent both the most comprehensive state of knowledge on the integration of GBA/GBA+ in evaluation given that the Department of Women and Gender Equality (WAGE, formerly Status of Women Canada [SWC]) elected to skip the administration of the annual GBA/GBA+ implementation survey in 2019.

Table 2 Variables by conceptual category

Awareness	Conceptual awareness of GBA+ Familiarity with GBA+ principles Identification of relevance of GBA+ for evaluation Identification of value of GBA+ for evaluation Identification of benefit of GBA+ for evaluation Alignment of evaluation and GBA+ Alignment of the assessment component between evaluation and GBA+ Products aligned
Policy Influence	Requirement to integrate GBA+ in evaluation noted Requirement to integrate GBA+ across policy cycle noted Belief in greater integration because of policy requirements Perception of reasonableness in adding to evaluation function responsibilities
Enabling Environment	Identification of supportive organizations Identification of supportive management
Tools and Training	Awareness of SWC/WAGE GBA/GBA+ tool Identification of SWC/WAGE GBA/GBA+ training received Identification of any GBA/GBA+ training received Awareness of GBA+/evaluation guidance materials Access to GBA+/evaluation guidance materials Engagement of Gender Champions Engagement of Gender Focal Points
Capacity	Comfort in application

	Perception of current sufficient capacity to integrate GBA+ in evaluation Perception of necessary financial resources to integrate GBA+ in evaluation Perception of access to internal expertise Perception of access to external expertise Access to tools to support integration of GBA+/evaluation Development of evaluation tools to support integration that are unique to organization Consistent integration
Decision-making	DM GBA+ consideration in evaluation DM GBA+ request for evaluation PM GBA+ consideration in evaluation PM GBA+ request for evaluation
Intersectionality	Program diversity element Different handling of program diversity element Systematic collection of GBA+ collection Concern that systematic collection of GBA+ information

Using SPSS, crosstabs with layers were undertaken. Layers were comprised of variables identified in the HoE survey and included “organizational size” and “government sector” to gain additional insight. A decision to not include the variables “number of evaluations conducted per year” and “capacity to integrate GBA” in layered crosstabs was predicated on self-identification by respondents; previous survey respondents had indicated respondent bias in overestimating functional knowledge and competencies relating to GBA+. Findings from these more nuanced crosstabs were used to explore and debunk multiple assumptions made at the onset of research. These assumptions covered all levels across the tailored ToC and are presented by the conceptual category below. Each of the conceptual categories generated between 44 and 100 pages of SPSS output which are available in electronic format upon request. Conceptual categorizations are

presented below according to their assigned level in the tailored ToC. Any counts presented are specific to these crosstabs.

Awareness and perceptions

It was assumed that all federal government organizations would meet all of the awareness dimensions. Layered crosstabs highlighted that awareness of GBA/GBA+ policy requirements were recognized by both Large Departments/Agencies (LDAs) and their Small Department/Agencies (SDAs) counterparts almost equally. Using sector, it was assumed that the Social Affairs (SA) sector would have a higher conceptual awareness of GBA/GBA+ because of the nature of the interventions in providing social assistance to diverse individuals. Crosstab results highlighted this was true, but surprisingly, the Economic Affairs (EA) portfolio also reported a high awareness of GBA/GBA+. EA respondents represented a much higher proportion of respondents than all others. It was also assumed that the International Affairs portfolio (IA) would have had a high awareness of GBA/GBA+ because of its ties to the government's international feminist assistance policy implemented as part of the liberal government's sweeping gender reforms but layered crosstabs reflect comparative awareness with other portfolio areas.

Policy requirements

It was also assumed that all federal government organizations would meet all of the policy requirement dimensions regardless of their size. Of direct relevance, 2 SDAs identified their unfamiliarity with the requirement to integrate GBA/+ in evaluation products and processes. More spectacularly, 6 LDAs identified their unfamiliarity with requirements to integrate GBA/+ in other policy cycle junctures. This significant proportion may be attributable to the timing of the *Policy on Results* (TBS, 2016) and subsequent guidance materials (TBS, 2017) which

highlights other junctures implicating the federal government evaluation function as they integrate GBA/+ across policy cycle junctures as the survey was launched just months after the release of the policy guidance's release. Assumptions on these two dimensions of policy were false. Respondents were also queried on their beliefs that greater integration would occur because of the policy. The assumption was first articulated that federal government organizations, particularly SDAs would agree. Approximately half of the total respondents LDAs (n=7) and SDAs (n=8) noted that the policy requirements would result in greater GBA/+ integration efforts for the federal government evaluation function. Layered crosstabs also highlighted that SDAs and LDAs alike (n=4 each) reported that greater integration of GBA/+ would not occur because of the policy. Insights provided as to why in other lines of evidence included: that GBA+ was already an organizational priority, that the mandate of the organization was not impacted by GBA/+, and direct services were not provided to individuals. It is unclear as to whether this assumption holds and would require further investigation. The assumption was made that most federal government organizations would identify that it was unreasonable to integrate GBA/+, however, all LDAs reported that it was reasonable, and only 2 SDAs reported that it was not. Examining assumptions at a sector level, it was believed that SA and GA would be well-represented across all of the awareness conditions. While these requirements were identified by these respondents, EA respondents identified a higher proportion as being aware of all awareness conditions which may be attributable to the reasons noted above.

Tools and training

Multiple dimensions are associated with the tools and training conceptual category. These dimensions begin with an awareness of SWC/WAGE's GBA+ tool and their free GBA+ training. Dimensions also include receipt of any GBA+ training such as that offered by Infonex (Findlay,

2019), access to evaluation specific tools to support integrating GBA/+ in evaluation, and lastly awareness and access to central agency guidance materials on integrating GBA+ in evaluation. It was assumed that all organizations regardless of size or sector would be aware of the federal government lead agency's tool and training which is available free of charge. This assumption was based on the federal government's long-standing commitment to gender equality, and the facility in accessing online training materials. This first assumption was found to be true as all LDAs or SDAs respondents were aware of the SWC/WAGE GBA+ tool. Surprisingly, not all organizations had taken the accompanying SWC/WAGE GBA+ training; 2 LDAs and 4 SDAs respondents had not taken this training. This finding is also evidenced in similar proportions when applied to any GBA+ training; 2 LDAs and 5 SDAs respondents had not partaken any GBA/+training.

The assumption was made as well that all federal government organizations would be equally aware of, and have access to central agency guidance materials on integrating GBA+ in evaluation despite organization size or sector. Despite these guidance materials being formalized in July 2019, various drafts were in circulation within the federal government evaluation community for consultation purposes. Respondents who had indicated not being aware of guidance materials included both SDAs (n=1) and LDAs (n=4) which is sufficient to prove the assumption false. Further, both LDAs and SDAs reported equally no access to these guidance materials. For future researchers, it would be reasonable to expect that all federal government organizations would report accessing guidance materials as they have been finalized and are available online for no cost.

It was likewise assumed that most federal government organizations would meet the majority of training dimensions regardless of size or sector. This assumption held when

examined by sector. Across all sectors, respondents noted awareness of SWC/WAGE's tool and comparable to organization size findings, not all sectors indicate having taken part in GBA+ training although the majority of respondents have (n=7/18). These same sectors identify approximately the same counts (EA=2, GA=2, SA=2) by sector as those having taken SWC/WAGE GBA+ training. These counts indicate the assumption as being false. The assumption regarding awareness and access to central agency guidance materials on integrating GBA+ in evaluation is more complex by sector. It was assumed that SA and GA would be more aware of, and have access to these guidance materials for the same reasons previously highlighted (assistance programs to individuals and alignment with government priorities). When examining more nuanced crosstabs it was found that a small proportion of respondents from these sectors had indicated not being aware of these materials (SA=1, GA=2). While a small proportion of total respondents, this highlights that the assumption holds only in some instances and further investigation is required. This small proportion increases when examining the dimension related to guidance materials access with the addition of 1 EA sector respondent. The final assumption regarding access to evaluation specific tools to integrate gender in evaluation was that all sectors would have had equal access to guidance, or how-to documents that would cover multiple dimensions of integration. As noted in the first several chapters, these documents are widely publicly available and are emerging with some frequency in the international domain. Findings from the detailed crosstabs illustrate that this assumption does not hold across sectors. The only sector that does not identify access to these tools is IA which resonates with the rationale provided above. All other sectors report not having access to these tools.

Enabling environments

There are four dimensions associated with this conceptual category that include supportive organizations and management, as well as engagement of Gender Champions (GCs) and Gender Focal Points (GFPs). As GBA/+ has been identified as a government priority including for the federal government evaluation function, it was assumed that support would be widespread across both organizational size and sector. Support was distinguished into two levels that included organization and management as it applied to the evaluation function. It was assumed that both support would be recognized almost universally at both the organizational, and management levels regardless of organizational size. This generally held as both LDAs and SDAs reported supportive organizations and supportive management. Only 1 SDA respondent indicated not supportive management. Assumptions are generally held in these two areas except for 1 EA respondent reporting a lack of support in each dimension. While interesting, it does not invalidate the assumption.

Additionally, the engagement of both Gender Focal Points (GFPs) and Gender Champions (GCs) were assumed to be critical components of an enabling environment given the investment of resources by Status of Women Canada/Women and Gender Equality into developing this gendered expertise. Responses highlighting the lack of engagement of these two resources were much higher by both organization size and to some degree consistent across all sectors. All sectors (SA, GA, EA, and IA) affairs indicated a lack of engagement with their GFPs representing almost half of total responses (n=11). The same assumption and findings were reported for the engagement of GCs. Consequently, these assumptions were found to be false. This suggests a large chasm between the evaluation and gendered expertise functions that was also noted previously although conversely two of the case studies. Reasons noted from previous chapters that could conceivably contribute include differential use of results language,

overextension of gendered expertise, the inability of federal government organizations to fill these positions.

Capacity

Several associated dimensions reside within this conceptual category. These include comfort in application, perceptions of current sufficient capacity, perception of necessary finances, access to both internal/external expertise and tools to support integration of GBA/+, as well as the development of evaluation tools to support integration that are unique to the organization and perceptions of consistent integration.

The assumption was made that SA portfolio areas would report higher levels of comfort in integrating GBA/+ for the same reasons noted above. Where differences emerged between assumptions and state of practice at the time, noted that more frequently than not, SDAs were more comfortable in applying GBA+ than LDAs. This may be because SDAs feel that because they are smaller organizations, that they not subject and do not experience the same degree of scrutiny that LDAs do. It was assumed that most organizations regardless of size report having sufficient capacity despite historical requirements to integrate GBA/+ in evaluation efforts. Findings were inconsistent with this assumption across both LDAs and SDAs although a greater proportion of respondents from SDAs note a lack of current sufficient capacity compared to their LDA counterparts. Breaking down elements of capacity, the assumption was made that most federal government organizations would not have necessary financial resources nor have established internal expertise at this stage relative to the requirement being articulated in the *Policy on Results* (TBS, 2016). Just under half of the total respondents identified necessary financial resources in place surprisingly. This was equal (n=4) across both LDAs and SDAs. This assumption was thereby declared false although with the caveat that respondents may not be

fully aware of the resource burden accompanying the integration of GBA/+ in evaluation efforts. In light of this, respondents also reported that just under half of the total respondents reported internal expertise. For the remainder of respondents (n=13), SDAs report a deficit in this area more than twice their LDA cohorts. It was assumed that both LDA and SDA respondents would be able to access external expertise and able to access tools to support integration efforts. Findings however suggest in both dimensions that almost equally both LDAs and SDAs do not have access to these resources. It was believed that if organizations did not have access to these resources, that they may perhaps then create their own which aligned with content previously referenced that identified a need for GBA/+ tools reflective of unique organizational environments and evaluation functions. Only a small proportion of both LDAs and SDAs have developed unique organizational tools thus nullifying this assumption.

While data were collected on perceived levels of related competencies, it is relevant to present findings in this section on capacity as they are directly related. It was anticipated that most federal government organizations would report the lowest capacity (on a scale of 5 options that included low, developing, established, intermediate and advanced). The majority of respondents identified a developing capacity, followed by established and intermediate levels. It was also assumed that those with established or intermediate capacities would report a higher comfort level in application which was found to be false as those with developing capacities reported a higher comfort in application. A potential rationale for this includes an openness to trying new approaches and methods with little fear of reprisal given relative established capacity levels. Briefly, a methodological application competence exists in the face of an early competence to identify conceptual issues or think evaluatively about GBA/GBA+. It was also assumed that those with intermediate or established capacities would report higher levels of

training received which was false as well. Contrarily, it is those with developing capacities that report receiving GBA+ training. This alludes to relevant experience being gained elsewhere.

Interestingly, EA portfolio respondents indicated the highest counts in these areas rather than those in SA portfolio respondents, followed by Governmental Affairs (GA) then IA. Given that GBA/GBA+ has been widely recognized as a federal government-wide priority, it was also assumed that GA would report the highest counts of GBA+ training; this did not hold as EA reported the highest counts of GBA+ training which lends credence to arguments forwarded that GBA/GBA+ essentially boils down to solid cost-effectiveness practice.

Decision-making

Conceptual categories related to GBA/+ decision-making included four dimensions. These four dimensions were grounded in research identifying that evaluation scoping within the federal government context is a negotiated process (Bourgeois & Whynot, 2018). Early evaluation scoping predominantly involves representatives from the program and evaluation functions defining the evaluation's focus including key evaluation questions, sub-questions, and related indicators. Scoping discussions would identify to what extent GBA/+ would be integrated in evaluation efforts, and at what level GBA/+ would be included in subsequent evaluation frameworks. Drafted evaluation frameworks are presented to Performance Measurement and Evaluation Committees (PMECs) which are comprised of executive level representatives from across the organization, for approval before any evaluation data collection efforts begin. Essentially, the decision-making dimensions associated with program management would capture working-level discussions and direction, and the dimension associated with organizational management would capture senior level discussions and direction.

Assumptions made focused on GBA/+ being highlighted at executive level discussions. This was premised on the perception that executives would want to ensure that the functions under their organizational jurisdiction would be aligned with federal government priorities. This assumption was true for LDAs where executive level consideration of GBA/+ was always a discussion point. This varied for SDAs however who reported an almost equal proportion of positive/negative responses of GBA/+ consideration. Consideration of GBA/+ does not directly translate to requests for GBA/+ integration in evaluation efforts which is reflected in the following dimension. The assumption was made that both LDAs and SDAs would subsequently request the integration of GBA/+ in all evaluation efforts. This did not hold; only three LDAs and SDAs each reported requests for GBA/+ integration. Delving down a level of seniority to the program manager level, assumptions were made that program managers would both consider and request GBA/+ integration in evaluation efforts. This assumption could not be verified; almost equal proportions of LDAs and SDAs responded that GBA/+ was not considered comparative to GBA/+ being considered. Further, in program managers' requests for GBA/+ integration in evaluation efforts, differences between LDAs and SDAs were almost equal in their positive and negative responses.

Examining the decision-making conceptual category by their government sector, the assumption that all sectors would be equally, highly represented by executive and program manager consideration and requests to integrate GBA/+ was made using the same rationale previously provided. Data supports the conclusion that this is untrue; 2 sectors' respondents (EA and GA) noted both negative and positive responses almost equally split amongst the two sectors regarding executive consideration of integrating GBA/+ in evaluation efforts. The remaining two sectors (IA and SA) report only positive responses. These latter responses lend credence to the

argument previously presented on the international feminist assistance policy as well as the direct provision of social assistance interventions to individuals. Further, when examining executive requests to integrate GBA/+ in evaluation efforts, only EA and SA respondents noted positive responses (n=3, n=1 respectively). Across all sectors, the mode response was two for the program manager requesting consideration of GBA/+ integration; only SA respondents indicated a “three”. Findings consistent with those presented above regarding program managers’ request to integrate GBA/+ in which negative responses from all sectors accounted for slightly over 50 percent of total responses. Appraisal by sector on this dimension of decision making was inconclusive with varied results rendering the assumptions false.

Intersectionality

This conceptual category was accompanied by 4 associated dimensions. These included having a diversity element embedded in the program/evaluand, whether attendant evaluation efforts were handled differently, whether GBA/+ information was collected systematically by the organization, and whether this collection was a concern. Regarding the first dimension, the assumption was made that only a small proportion of federal government organizations would have a specific diversity element embedded in their programs. This was rationalized despite new policy cycle requirements requiring the integration of GBA+ which were only recently added. Until 2017-2018, federal government interventions were not required to identify differential impacts in their design phase. In examining detailed crosstabs by organizational size, it appears that almost half of the responses identified having a specific diversity element (LDAs n=3, SDA n=7) in programs/evaluands which nullifies the assumption. Surprisingly, many more SDAs appear to have diversity elements in their programs/evaluands than their LDA counterparts. Continuing with the associated dimension as to whether the evaluation was handled differently

as a result of the inclusion of this element of diversity, the assumption was made that all organizations would not handle it differently. According to respondent accounts, only minimal identifications from LDAs (n=2) and SDAs (n=1) enacted differential handling. While a relatively smaller count by number, this represents almost a quarter of total responses and the subsequent assumption was interpreted to be false. One of the most important dimensions associated with this conceptual category is the systematic organizational collection of GBA/+ information by the performance measurement function. All international guidance documents suggest that disaggregated data must be available to undertake intersectional analysis (UNEG, 2015; UNEG, 2018; UNWomen 2014). It was assumed that LDAs would report more consistently than their SDA counterparts, the systematic organizational collection of GBA/+ related data and if they did not, this would be a concern. This was because it was believed that performance measurement functions would have an increased capacity to do so, with accompanying resources. When regarding the nuanced crosstabs, it was found almost equally across organizational size that almost half of LDAs and SDAs do not systematically collect GBA/+ information. Surprisingly, just less than half of LDAs and SDAs respondents identified this as a concern. This will be a significant challenge given the consistent iteration of calls for disaggregated data articulated in both domestic and international guidance materials. This challenge also speaks to the need for greater coordination between the evaluation and performance measurement function on all aspects of GBA/+ information.

Finally, when examining the conceptual category related to intersectionality and its associated dimensions by their government sector, many of the same sectoral assumptions were made as have been articulated in sections above. Sector differences concerning diversity elements embedded in programs/evaluations were assumed to be few. As with the findings

presented above, this was not the case. All sectors' responses noted programs/evaluands with diversity, representing almost half the total responses. These accounts nullify the assumption made. For differential handling of evaluation efforts to respond to diversity within the program/evaluand, it was assumed that no variance by sector would occur. Two sectors (EA and IA) identified differential handling within evaluation efforts. While a relatively small proportion of total responses, this suggests a potential area for future research. On the third dimension of systematic organizational collection of GBA/+ information by the performance measurement function, it was assumed similarly to organizational size, that some differences would be present. Only approximately one-third of the total respondents indicated positively consistent GBA+ data collection. By sector, EA, GA, and SA report positively systematic GBA+ data collection. This leaves approximately two-thirds reporting non-systematic GBA+ data collection with varying accounts across all sectors.

In summary, significant differences emerged between assumptions and state of practice on integrating GBA+ in evaluation. More frequently than not, SDAs were more comfortable in applying GBA+ than LDAs. Potential contributing factors highlight that both LDAs and SDAs had not taken the SWC/WAGE GBA+ training in equal proportions despite being aware of its existence. Additional insight is provided in referencing that despite SDAs reporting greater comfort levels, a more significant proportion of SDAs highlight not having taken this training. Another assumption in the area of organizational size was premised on the belief that LDAs would systematically collect GBA+ information because of a greater number of resources available to them. This did not hold; the opposite was reported. SDAs (or Micro Departments/Agencies [MDA]) consistently reported systematic collection of GBA+ information than their LDA counterparts. A final assumption worth noting presumed that LDAs might more

consistently engage with their gendered expertise (Gender Champions or Gender Focal Points) than SDAs. The inverse was reported with SDAs highlighting more frequent engagement with their Gender Champions; Gender Focal Points were reported with equal counts. The rationale driving SDA increases over their LDA counterparts might stem from their flexibility to nimbly navigate government and organizational priorities as they do not have the processes and silos to navigate that LDAs do.

Overall, it was assumed that the Social Affairs (SA) sector would have a higher conceptual awareness of GBA/GBA+ because of the nature of the interventions falling within their intervention areas in which individuals are provided social program assistance. Crosstab results highlighted this was true, but surprisingly, the Economic Affairs (EA) portfolio also reported a high awareness of GBA/GBA+. It was also assumed that the International Affairs portfolio would have had a high awareness of GBA/GBA+ because of its ties to the government's international feminist assistance policy but crosstabs reflect comparative awareness with other portfolio areas. The assumption was also made that SA portfolio areas would report higher levels of comfort in GBA/GBA+ training and comfort in application for the same reasons noted above. Interestingly, EA portfolio respondents indicated the highest counts in these areas rather than those in SA portfolio respondents, followed by Governmental Affairs (GA) then IA. Given that GBA/GBA+ has been widely recognized as a federal government-wide priority, it was also assumed that GA would report the highest counts of GBA+ training; this did not hold as EA reported the highest counts of GBA+ training which lends credence to arguments forwarded that GBA/GBA+ essentially boils down to solid cost-effectiveness practice.

Finally, it was anticipated that most federal government organizations would report the lowest capacity (on a scale of 5 options that included low, developing, established, intermediate

and advanced). The majority of respondents identified a developing capacity, followed by established and intermediate levels. It was also assumed that those with established or intermediate capacities would report a higher comfort level in application which was found to be false as those with developing capacities reported a higher comfort in application. A potential rationale for this includes an openness to trying new approaches and methods with little fear of reprisal. Briefly, a methodological application competence exists to identify conceptual issues or think evaluatively about GBA/GBA+. It was also assumed that those with intermediate or established capacities would report higher levels of training received which was false as well. Contrarily, it is those with developing capacities that report receiving GBA+ training. This alludes to relevant experience being gained elsewhere.

While the summary provides a few insights, it also creates many questions. Varying organizational results complicate analysis but one of the key research questions includes what the various kinds of evaluation uses stemming from the integration of GBA+ in the federal government evaluation function have resulted. Below, instrument, symbolic, process, and conceptual uses are discussed.

Gender-based Analysis Plus (GBA+) instrumental use

Producing GBA+ information is important; equally important is stakeholders' use of GBA+ information. Recent integration efforts are laudable, however, concerns specific to the federal government evaluation function should be explored further. Instrumental evaluation use is the most directly aligned with the *Policy on Results* (TBS, 2016) objective. The focus resides on improving the results achieved by the Canadian government. Definitions of instrumental use comprise two components that include evaluation findings being used to directly inform decision-making, and findings being linked to resulting action. Instrumental evaluation use then

occurs when decision-makers use evaluation findings to modify the object of the evaluation in some way (Greene, 1988; Shulha & Cousins, 1997). Evaluation recommendations and management responses are a key dimension of making evaluation reports public; this is one way in which GBA+ instrumental use can be examined.

Unlike the United Nations (UN) body of organizations and like-minded efforts in the WorldBank, the Canadian federal government has no requirements to report on an annual basis how, and to what extent associated organizations have integrated gender in their evaluation efforts. Reviews of these reporting requirements have identified that “just over half of UN entities are at least “meeting” requirements, which leaves almost half of UN entities not meeting requirements (UNEG, 2016, p.4¹). Within the Canadian federal government, central agencies are now contemplating how best to achieve system-wide reporting on the impacts of federal organization’s efforts regarding gender and diversity. This was alluded to in Budget 2019 with the allocation of “1.5 million over 5 years, starting in 2019-20 to the Treasury Board Secretariat to work with departments receiving Budget 2019 funding to ensure robust administrative data collection and reporting practices with respect to gender-based analysis plus (GBA+) information for all initiatives” (Budget, 2019). The important recognition of an accountability mechanism to monitor EDI is significant, but as previously stated, EDI is not GBA+.

In participating case study organizations where a specific evaluation (n=4) was tracked, instrumental use was examined. In all 4 organizations, it was found all GBA+ dimensions were represented to some extent. This was dependent upon the evaluation foci. GBA+ was specifically referenced in 2 of the 4 individual evaluations which could reflect the extent of the organization’s internalization of GBA+ as a federal government, and by extension, the evaluation functions’ priority. In these 2 instances, programs’ lack of performance measurement information

to support undertaking GBA+ were identified as challenges. They were urged to find alternative sources of information to ensure the availability of necessary information. While this is a promising finding, only in 1 instance did GBA+ emerge within the context of a specific recommendation. This recommendation pointedly referred to the program's need to consult with individuals from diverse backgrounds in the program design phase as part of the Ts&Cs renewal process. In another participating case study organization, the identification of the availability of GBA+ information and its reference in the evaluation report directly led to a related follow-up study investigating a specific aspect of GBA+ both in-depth, and in context at the request of organization senior-decision makers. Discussed above are instances in which micro-level instrumental evaluation use (TBS, 2012) are identified. Macro-level instrumental evaluation use has also been identified which includes consultation on TB Subs and MCs, as well as evaluation function involvement in Parliamentary reporting. In all participating case study organizations, HoE indicated that they were consulted during the preparation of all these documents however, there was disagreement and a lack of knowledge as to whether their input was considered in the final drafts of the documents noted above.

At the conclusion of this research, less instrumental evaluation use was evidenced than expected. This finding is echoed in the academic literature which presents an organizational case study that explores the relationship between integrating GBA+ and use (Johnson & Momani, 2019). These authors concluded that very little instrumental use had occurred by decision-makers resulting from GBA+ integration. GBA+'s design intention was to ensure that GBA/+ provides evidence to support decision-making that would improve/achieve EDI and when integrated in evaluation, that both an improved quality and quantity of information to support decision-making would result. Stated another way, very little information was produced in ways that would enable

impact level changes in EDI. Compounding this lack of micro-level GBA+ instrumental use, research highlighted that organizational focus consistently emphasized the first two elements; the “E” and “D” with indicators representing numbers of, barriers experienced, etc. rather than focusing on the “I”. That EDI and GBA/+ are used interchangeably exacerbates this issue. There appears to be a widespread misunderstanding of the relationship between GBA/+ and EDI.

Gender-based Analysis Plus (GBA+) symbolic use

Another potential evaluation use is symbolic use (Alkin & Taut, 2003; Greene, 1988; Leviton & Hughes, 1981; Weiss, 2016). In some instances, this has been referred to as tokenistic, or rhetorical support for an evaluation process or findings but with little intention to take the process or findings seriously. Studies into this use have become more prevalent, particularly because of its significance in policy discourse. This poses a very real concern for the Canadian federal government evaluation function as the function struggles to find its balancing point on the accountability-learning teeter-totter. Federal government evaluations are required by legislation but multiple resource limitations face the function.

Echoing Marra’s (2003) findings on symbolic evaluation use, several interviewees referenced the policy requirement to examine both the extent of the program’s implementation of GBA+ implementation and integrating GBA+ in evaluations. Cautionary notes were identified included; if GBA+ related recommendations didn’t appear in the final product instrumental use would be difficult to track, and fear of superficial or rote GBA+ application would not ultimately result in beneficial use. The Government of Canada referenced this latter dimension of GBA+ application in its latest Budget (Finance Canada, 2019) citing “...the quality and application of GBA varied. In some cases, GBA+ done by departments was cursory. In other cases, insufficient analyses could be performed due to a lack of data, particularly with intersecting identities”

(p.279). Simultaneously, interviewees also endorsed the inclusion of GBA+ in evaluation policy as a legitimizing factor for the embodiment of GBA/+ in evaluation. The study of GBA+ implementation across mandated requirement areas including budget proposals may be of interest to future researchers. Accounts from the Canadian evaluation grey literature covering the policy transition period extending from 2015-2018 highlighted an expected increase of (S)GBA+ which was not realized (Borys, Gauthier, Kishchuk, Roy, 2019). Further, where (S)GBA+ was present in the sample of reviewed evaluation reports, limitations and/or constraints were noted (Borys, Gauthier, Kishchuk, Roy, 2019) rather than findings being presented from any level of intersectional analysis. This acknowledgment of, but failure to act, can be imagined as a form of symbolic evaluation use.

Peck & Gorzalski (2009) identify that symbolic evaluation use is employed to convince others of a political position that is highly intertwined with organizational reputation and/or status (Alkin, 2003; Shulha & Cousins, 1996). When acknowledging the requirement for GBA/+ in evaluation, there is a strategic purpose underscoring its reference. In doing so, there is a formal recognition of federal government GBA/+ priorities but positions either the program or evaluation function as being unable to address these priorities. Reasons provided included: performance measurement information being unavailable at individual levels, related privacy concerns and protections associated with the collection, storage, and use of individual-level information; as well as the requirement to build capacities within the evaluation function to integrate GBA/+ in evaluation products and processes within existing evaluation processes (such as DEP). In addition, although this was not explicitly articulated during data collection, GBA+ was also built into the performance agreements of senior officials in some of the federal government organizations. Thereby in acknowledging that GBA/+ was recognized as a federal

government priority in evaluation reports, the requirements of performance agreements could be considered as met.

Gender-based Analysis Plus (GBA+) process use

Beyond instrument and symbolic GBA+ uses, significant evidence was found on GBA+ process use. This kind of use focuses broadly on “the utility to stakeholders of being involved in the planning and implementation of an evaluation” (Forss et al., 2002, p.30). Process use may be exhibited at both individual and/or organizational levels (Patton, 2008). Patton defines this distinction as “individual changes in thinking, attitudes, and behavior, and program or organizational changes in procedures and culture that occur among those involved (2008, p.155). The integration of GBA+ across the policy cycle diagram presented in earlier chapters of the thesis highlights the extent to which GBA/+ is integrated into various policy development phases but this does not reflect the integration of GBA/+ across evaluation phases including planning/design, data collection, analysis, and reporting levels. Various kinds of process use have been identified by evaluation theorists (Forss, Rebien & Carlsson, 2002; Patton, 2008). A consensus has been established that generally highlights the following kinds of process use which evaluation stakeholders may engage in integrating GBA/GBA+ that includes the following learning to learn, developing networks, creating shared understanding, strengthening the project; and boosting morale.

Demonstrations of GBA/+ process use amongst case study organizations included workshops/information sessions held at varying intervals and durations including both individual events and a series of events held over several months. Workshops/sessions focused on building capacity to integrate GBA/+ both within the evaluation function, and more broadly within the organization involving functions supporting evaluation such as policy, performance

measurement, communications, and gendered expertise. Capacity building workshops/sessions were held in two formats that included intentionally focused on the various evaluation phases noted above, and others focusing on evaluation issues where the intersection between the evaluation and performance measurement functions are critical. In the latter workshops/sessions representatives from the evaluation, performance measurement, policy, and gendered/diversity expertise were present. In holding these workshops/sessions, common understandings were created that were unique to the organization, and learning occurred at the individual, at functional levels, and between functions. In addition, for case study organizations focused on building capacity at the functional level, other representatives from across organizations were also present and included: communications, privacy, and research functions. At times, these individuals were involved to provide organizational context, and/or GBA/+ details that evaluation personnel may have been unaware of. It appeared that more intensive learning occurred for evaluation personnel when the cross-section of stakeholders was narrowed because discussions were more specific, and intensive relative to evaluation processes and products. Learning also occurred when workshops/sessions focused on specific evaluations. When discussing specific evaluations, only evaluation personnel were involved.

Throughout the research period, some members across the federal government's evaluation functions demonstrated the acquisition of a capacity to think "evaluatively", a key component of process use (Patton, 2008) in integrating GBA/+. The capacity to think evaluatively as applied to GBA/+, manifests with a facility to conceptualize grasp how EDI issues may manifest in the evaluand, may apply to evaluation efforts, and be adapted for the examination across the various evaluation phases. Thinking evaluatively regarding GBA/+ serves as an important precursor to integrating GBA/+ in evaluation; if the ability to recognize existing potential barriers to access

and participation are identifiable which can be explored and potentially widen the door to identifying and addressing the perpetuation of various forms of discrimination in subsequent evaluation efforts. There is a moral component associated with the application of thinking evaluatively about GBA+ that necessarily requires the assumption that all forms of equity are desired. If equity is not the desired state then equity challenges cannot be raised nor addressed which can then contribute to the perpetuation of systemic inequality.

The proficiency to think evaluatively regarding the integration of GBA/GBA+ in evaluation was facilitated by previous instances in which inequality had emerged. Instances in which inequality emerged were identified by various sources including potential program beneficiaries who might have experienced program access and/or participation barriers, representatives from the program in possession of strong performance measurement functions and monitored inequalities in program activities, and by evaluation personnel. For example, highlighting disproportionate funding in favour of men compared to early-career women scientists in the health domain proved beneficial to evaluators in other domains because of marked similarities in funding programs. Where applying this evaluative thinking to GBA/+ to conceptually identify the EDI issue was found to be more challenging lay in thematic areas that had little or no experience applying such concepts. This is in part attributable to the broad thematic reach which could potentially breach distinguished organizational mandates, and responsibility areas were less clearly defined. Interestingly, this ability to think evaluatively regarding GBA/+ was less present in programs focusing on EDI issues. Interviewees and survey data consistently highlighted that GBA/+ was either unnecessary or less necessary, as EDI issues were already being addressed. Interestingly, these case study examples did not consistently

collect performance measurement data at the individual level which contrasts sharply with suggestions in formal guidance materials.

Federal government organizations often possess informal GBA+ expertise within their evaluation units. The identification of this individual as a GBA+ point person stems from three factors that include; personal interests, dedication to equity, and this individual generally holds some seniority within the organization's evaluation function. One of the most significant demonstrations of GBA+ process use amongst case study organizations was the convening of organizations' evaluation functions to learn together. Several case study organizations set aside time to ensure comprehensive discussions as to how the function would tailor evaluation efforts; this involved discussing"

- Developing systematic integrative approaches that addressed standardizing key evaluation questions, indicators and questions to be included in each line of evaluation evidence;
- Conducting post-mortems on previous and current evaluation efforts to support improvements in future integrative evaluation efforts; and
- Engaging expertise to provide evaluation function personnel with supported tailoring of integrative evaluation efforts.

Many of these GBA/+ process use discussions were integrated into or led to the creation of federal government evaluation function networks in which emerging practices were shared amongst representatives of federal government evaluation functions. Multiple local evaluation networks including the Canadian Evaluation Society's National Capital Chapter (CES-NCC); the Lunch & Learn for Evaluators (L&L), and the Planning and Performance Exchange (PPX)all hosted multiple sessions from 2017-2020 focusing on integrating GBA/GBA+ in evaluation.

These networks are driven predominantly by the needs of federal government evaluators. In addition, in May 2020 CES National hosted a WebEx session on *Integrating GBA+ in Evaluation* which represented the highest attended webinar in their history. This researcher has either presented or been involved in the planning of all these network events. The support of these various networks in exploring the integrating of GBA/GBA+ ensures a rich learning environment and support beyond one's organizational function.

Gender-based Analysis Plus (GBA+) conceptual use

Conceptual evaluation use occurs when individuals, as a result of evaluation efforts and/or influences, think differently about the object of the evaluation (Rich, 1977). Emphasis is placed on changes in levels of knowledge, understanding, and attitude (Johnson, K., Greenesid, S., King, J.A., Lawrenz, F. & Volkov, B, 2009; Peck & Gorzalski, 2009). Unlike instrumental use, no action or decision need result but rather, evaluation exposure helps create a better understanding of the program, or its beneficiaries (Rich, 1977; Weiss, 2016). For example, when evaluation processes assist staff to understand the program (or its diverse beneficiaries) in a new way conceptual evaluation use is said to have occurred.

Guidance materials offered by central agencies were designed to address conceptual GBA+ use. One of the main critiques of the published guidance materials is the lack of detail on what constitutes "intersectional". The guidance materials state that "the 'plus in GBA+ acknowledges that GBA goes beyond biological (sex) and socio-cultural (gender) differences. We all have multiple identity factors that intersect to make us who we are" (TBS, 2019). No additional detail is provided. Guidance materials do not explicitly reference intersectional analysis, nor provide any detail on what intersectional analysis might involve. As introductory materials, this should be at the forefront of guidance materials intended for a function that had

not been previously targeted for GBA/+ capacity building efforts. Historically, program and policy areas were prioritized for GBA/+ capacity building efforts, and resources were allocated accordingly.

Additionally, guidance materials do not highlight that sex and/or gender may not be the most important identity variables for any federal government policy, program, process, or intervention at the time of evaluation. Recent academic offerings exploring inequality note that some identity variables are more important than others at various stages of an individual's life. Ragin and Fiss (2017) note that poverty was found to be more important than race (Ragin & Fiss, 2017) in students' access to academic programs. This aligns with WAGE's recent changes to GBA+ guidance that suggest depending on an organization's mandate, gender/sex may not be the most important identity variable (WAGE, 2018) to consider. The failure to introduce these issues in guidance materials could potentially caution unintended detrimental outcomes.

Similar to other policy cycle junctures and discussed in previous chapters of this thesis, guidance materials also provide a caveat of "where relevant" in guidance materials. This caveat could potentially serve as an "out" for those who are unable to think evaluatively about their application of GBA+ to the evaluand. Despite these roadblocks, this is a significant step towards improving GBA+ conceptual use; materials address a void, serve to legitimize evaluator efforts to improve the integration efforts, and may ultimately improve future EDI outcomes.

To conclude, less instrumental GBA+ use than anticipated resulted but promisingly, greater process GBA+ was identified across the federal government's evaluation function. While greater process GBA+ use is a necessary precursor for increased instrumental use, a risk exists GBA+ is unable to prove its value to decision-makers and without sufficient monitoring of the extent of GBA+ integration, a risk exists that organizational and functional leadership may

withdraw their support and may only implement tokenistic applications which would ultimately continue to perpetuate existing systemic discriminations.

Chapter 8 – Dear Betty: Everyone has had dessert. What’s next?

The crafting of the final chapter fell during an unusual period of Canadian history. In March 2020, not only were Canadians mandated to socially distance themselves in light of a global pandemic but several months later in May/June 2020, protests erupted across North America in support of anti-black racism following the murder of George Floyd. The latter represented a critical breaking point in a long history of inappropriate policing against marginalized individuals, most notably Black and People of Colour (BIPOC). These two examples represent the health and justice systems that comprise only two of the institutions that constitute the Canadian federal government potentially perpetuating systemic discrimination. The common denominator amongst the current background involves the disproportionate loss of life experienced by marginalized individuals. These two systems represent “low-hanging” fruit in GBA/GBA+ application in which differences based on dimensions of personal identity are facilitated easily.

Returning to the literature

These low-hanging fruits are reflected in the academic literature. A return to the academic literature, focusing on Canadian content produced within the last 4 years using GBA/GBA+ as defining search terminology reflects numerous offerings within the health system. Among these health-related high-yield/low-risk applications of GBA/GBA+ are those highlighting drug use/substance abuse (Greaves, 2020) including illicit opioids (Cheng, Nosova, Sall, Hogg, Hayashi & Debeck; 2019; in the context of supervised injection sites (Kennedy, Hayashi, Milloy, Boyd, Wood & Kerr, 2020); pharmacology (McCarthy, Milne, Waite, Cooke, Cook, Chang & Sproule, 2017); psychosis (Haarmans, McKensie, Kidd & Bentall, 2018); palliative care (Sutherland, Ward-Griffin, McWilliam & Stajduhar, 2017; Sutherland, Ward-Griffin, McWilliam

& Stajduhar, 2016), health research (Cooke, Waite, Cook, Milne, Chang, McCarthy, & Sproule, 2017; Page, Dorais, Bilkis, Katz, Samb, Vanasse, Bartlett, Buckeridge, Hudon, Lamarche, Porney, Cloutier 2018), smoking (Bougie & Kohen, 2018); physical activity (Coen, Subedi, & Rosenberg, 2016); sleep disturbance (Conklin, Yao, & Richardson, 2019); domestic violence (Ahmadzie, Stewart & Sethi, 2016); cardiac/heart studies (Hoang-Kim, Parpia, Freitas, Austin, Ross, Wijesundera, Tu, Mak, Farkouh, Sun, Schull, Mason, Lee & Rochon, 2020) and workplace violence ((Lanthier, Bielecky & Smith, 2018).

Encouragingly, a smattering of GBA+ method-related offerings were also identified that included reflections on positioning gendered expertise within bureaucratic settings (Scala & Paterson, 2017); challenges and opportunities of GBA+ as an intersectional method (Hankivsy & Mussel, 2018); the intersection of GBA+ and policy (Dobrowsky, Macdonal, Raney, Collier, Dufour, Scala & Paterson, 2017; Findlay, 2019) and GBA+ case studies specific to the federal government (Drabble & McInnis, 2017; Johnson & Momani, 2019; Maillé, 2018).

Accompanying movement into unfamiliar terrain experiences of being uncomfortable were cited; this experience was embedded in one of the case studies (Johnson & Momani, 2019) in which the Department of National Defence's experiences in implementing GBA+ in a defence setting were grounded against a backdrop of historical gender discrimination. It is important to identify this backdrop because the practice implications that acknowledge wrongdoing based on discrimination may serve as an impediment to others in their implementation efforts. This nudge into unfamiliar terrain and uncomfortable conversations is ultimately grounded in existing value systems and individual roles within institutions challenging the status quo. Evaluation as a profession has grappled with how to measure and report on changing social norms which are an attendant dimension of improving systemic discrimination by way of culture shifts. While

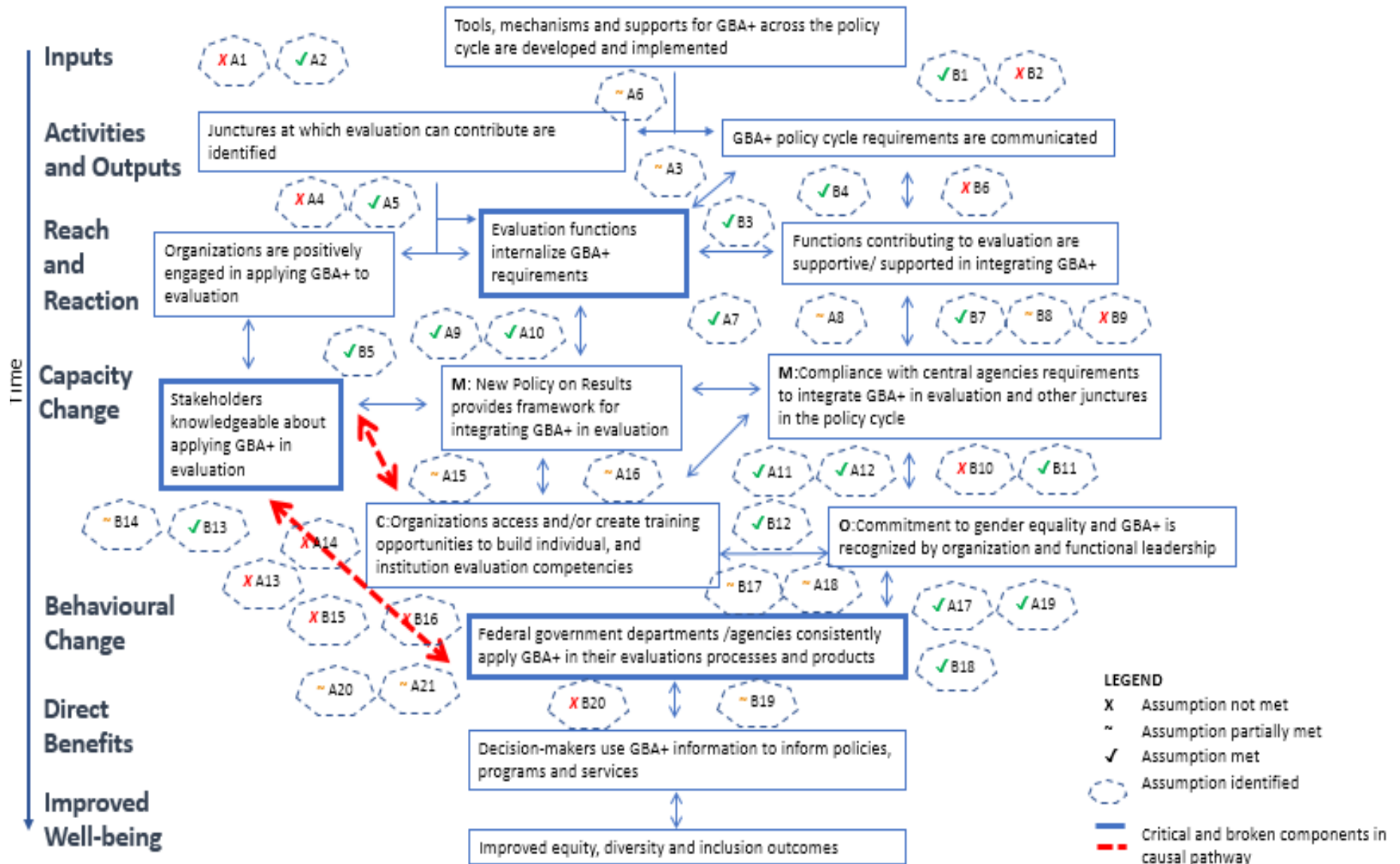
promising insights are emerging, challenges remain. In addition, promising offerings were identified in other domains including legal (McDonell, 2018) and climate/environmental studies (Chalifour, 2017; Rochette, 2016). Identified offerings in this paragraph are promising for their indication of expansion beyond the low-hanging GBA/GBA+ fruit. This is in part because these are domains that have not traditionally explored how dimensions of personal identity impact and are impacted by federal government interventions. The legitimatization of GBA+ beyond its historical feminist roots as reflected in academic journals is a powerful motivator for both internal and external stakeholders. It becomes more than “a feminist thing advanced by SWC” (now WAGE) (personal communication, 2018). When this research was initiated, a review of the academic literature using the same terminology returned less than 20 articles; the most recent results yielded a return of +50 resources including articles, books, and dedicated theses/dissertations. The inquiry into GBA/GBA+ has made inroads with its broadened authorship, discipline, and implementation efforts. One remaining future challenge includes potential offerings reflecting on efforts by federal government functional communities, particularly by functions that capture ex-post results.

Final reflections

For future operationalization of GBA+ across the policy cycle, within individual organizations, and across government functions including evaluation, there is value in returning to the tailored ToC that guided this research. While various findings were presented across previous thesis chapters, Figure 8: Updated Theory of Change (ToC) for Integrating Gender-based Analysis Plus (GBA+) in the Federal Government Evaluation Function presented below summarizes the comprehensive update of the causal impact pathway inclusive of the entire results chain and

accompanying assumptions. In this section of the final thesis chapter, the original tailored ToC is reflected upon and identified concerns surrounding assumptions pondered with implications for future research identified. Uniquely to the federal government's evaluation function, the graphic provides an updated status for each of the assumptions based on findings from this research using the following symbols: "x" representing did not meet assumptions; "~" for partially met assumptions, and; "✓" denotes having met assumptions.

Figure 8 Updated tailored theory of change (ToC) for integrating Gender-based Analysis Plus (GBA+) in evaluation



At the early stages of the causal impact pathway, assumptions were predominantly met. Two glaring deficits emerge however that will ultimately affect the achievement of decision-makers' use of GBA+ related information stemming from evaluation processes and products. Key challenges remain concerning knowledge of GBA+ by evaluators, and knowledge of evaluation by gendered expertise. Specifically, there is a lack of acknowledgment of the feminist principles regarding the production of knowledge which is reflected in the application of GBA+ within the evaluation function. Simultaneously, gendered expertise's unfamiliarity with the evaluation function, including differentiating between the performance measurement and evaluation function themselves, is problematic. Despite assumptions related to awareness, value, and benefits of integration efforts, the implications of not having critical assumptions being achieved so early in the results chain signals concern for the successful achievement of subsequent results chain phases and achieving the ultimate objective. Consequently, the reciprocating lack of knowledge impedes progress not only towards subsequent results chain phases but is potentially detrimental to improved EDI outcomes as well.

It is important to acknowledge political, organizational, and functional leadership and by extension, policy requirements. The integration of GBA+ across the federal government policy cycle has advanced more significantly under current political leadership within a 5-year timeframe despite Canada's government commitment stemming from the 1995 Beijing Declaration. This suggests that leadership is important, particularly political will is even more so. To a significant extent, federal government evaluation functions have internalized GBA+ requirements but lack the requisite knowledge to systematically integrate GBA+ efforts in evaluation processes and products. Sectors appearing to focus on EDI, rather than GBA+, have produced findings flowing from their evaluation processes and products that decision-makers can

use to improve EDI outcomes. It was previously noted that policy is an inherently messy process; the same rhetoric applies to GBA+; further, when integrating these two inherently wicked policy problems within unique organizational and evaluation contexts, leading these efforts are not for the faint of heart. To be clear, these changes are not intended to lead to transformative outcomes per se as originally envisioned with the adoption of GBA+ as a means to first gender mainstreaming, and subsequently mainstreaming for all equities. Perhaps equity advocates may find solace in this third wave of GBA+ implementation, that for the first time, federal government interventions have acknowledged the heterogeneous nature of their program beneficiaries and must respond accordingly in reducing barriers to access, and participation in various government programs. Repeated demands for illustrative examples within the evaluation function echoed loudly throughout this research; encouragingly, several case study organizations were able to provide sectoral relevant examples that facilitated uptake in other federal government organizations. Examples such as these however could be characterized as an anomaly rather than a common occurrence.

Building the capacity to apply GBA+ in evaluation processes and products begins with unique organizational, program, and evaluation knowledge. Revisiting the academic literature reflects organizational (sector) and program efforts to build GBA+ but until this research, integrating GBA+ in the federal government evaluation function was non-existent. Specific guidance materials were available only in draft format until 2019 and additionally, are missing significant GBA+ elements such as how to develop recommendations in consultation with programs that will support the achievement of EDI outcomes. Consequently, several evaluation functions have demonstrated initiative in developing guidance and training materials unique to their organizational context. How effective these materials are will require future research to

regard evaluation reports and accompanying Management Response and Action Plans (MRAPs). This research contributes in a novel manner; it outlines demonstrated the earliest attempts offered by the federal government evaluation function. Research offers insights as to how GBA+ may be directly and indirectly integrated against core issues, in evaluation frameworks, in lines of evidence, and across evaluation phases. It is evident that integration for evaluation requires reflection at evaluation initiation; integrating GBA+ must be planned for which requires the coordinated engagement of evaluation supporting functions such as programs, performance measurement, policy, research, etc. Too often, however, larger organizations face barriers for coordinated engagement to ensure that evaluation has the necessary data; smaller departments appear to be able to respond more nimbly to coordination requirements. This research also highlights longstanding relationship tensions between the performance measurement and evaluation functions. In this context, it is more important that the evaluation function can create GBA+ information if envisioned during the planning phase rather than relying on programs and the performance measurement function to procure this information. The possible role played by performance measurement has been noticeably absent in the integration of GBA+ across the entire policy cycle with evaluation unfairly shouldering disproportionate responsibilities.

Quality criteria will play an important role for reliable and meaningful GBA+ implementation; currently, standards do not exist that exemplify quality criteria that are associated with the confidence of decision-makers using GBA+ information provided by the evaluation function. Recent budget documents (such as the 2020 fall economic update) footnote quality GBA+ as timing dependent relevant to the status of program development rather than focusing on whether GBA+ content is effective. Future researchers would be well situated to investigate the GBA+ quality dimension further with the existing emphasis on being transparent

and accountable by ensuring this information publicly available. This research would also contribute to broader ToC gender mainstreaming efforts and strengthen the tailored ToC used by this research. Oftentimes in ToC, targets and benchmarks are established as a component of program theory implementation and monitoring. As this was the first detailed foray into the integration of GBA+ and the federal government evaluation function, it was unrealistic to set benchmarks and targets. The implications of not doing so are addressed in future directions for related research and are highlighted in detail in the following section of this thesis chapter.

Space for future Qualitative Comparative Analysis (QCA) efforts

How each federal government organization integrates GBA+ is uniquely dependent upon multiple factors explored and presented in this research. These organizational variances provide rich fodder for future explorations of causal complexity. In particular, Qualitative Comparative Analysis (QCA) was noted as one feasible approach to exploring causal complexity. This method is grounded in configurational complexities using algebraic Boolean reductions; QCA can be adopted as either research design or as an analytical process. The value in applying QCA to findings from this research would provide insight beyond the question “and so what” for more definitive outcomes.

Prerequisites supporting QCA include a small dataset of more than 8 cases ($n > 8$), formal logic which was represented by the tailored ToC, the presence of causal complexity (organizational variances), and data to support condition calibration (HoE survey and case study findings). Amendments to original research questions to align with a QCA application would resemble “What combination of leadership, awareness, requirements, training, tools/supports, and methodological capacities are present that lead to the consistent integration of GBA/GBA+

in federal government evaluation function”? There are six steps involved in undertaking a QCA (Kahwati & Lane, 2020) include the following:

1. Transform a data matrix of set membership values into a truth table
2. Management of contradictory truth table rows
3. Revisit data to manage contradictions
4. Inspect the truth table for potential issues
5. Conduct analysis of necessary conditions and combination of conditions
6. Conduct preliminary sufficiency analysis of truth table

In QCA, calibration of set membership is defined in one of three ways that include Crisp-Set (csQCA), Fuzzy Set (fsQCA) or Multi-value set (mvQCA). Set calibration is the process of “converting raw case data for the conditions and outcome into numeric set membership values [SMV] that represent each case’s membership in each condition and outcome set specified in the analysis” (Kahwati & Lane, 2020, p.66). Possible SMV ranges in binary expression from 0 (nonmembership) to 1 (full membership). In csQCA, cases may only assume either nonmembership or full membership SMV whereas fsQCA, cases are expressed on a continuum between nonmembership and full membership; mvQCA employs both.

Two major limitations existed in applying QCA as an analytical tool in the context of this research; conceptually, the constantly changing GBA+ policy landscape, and more practically, this researcher’s unfamiliarity with the QCA application with R software. To address the conceptual concern, a mvQCA would have to be applied. This would require updating the HoE survey to incorporate scales that reflect the extent of existence rather than relying on scales representing the presence of a particular condition. Admittedly, I had initiated early csQCA efforts using Tosmana software based on its user-intuitive appeal but realized after attending a

QCA paper development session where methods experts had highlighted my possible QCA application's shortcomings. Professional advice had included switching from Tosmana to R software and paring down original conditions included in the analysis. This research is worth pursuing as novice QCA applications provided some interesting elimination of conditions after necessity and sufficient analysis were conducted. While statistically significant, removals included "policy requirements" were defined as trivial conditions, and the "systematic collection of disaggregated performance measurement information by identity dimension" because contradictory conditions cannot exist in both the outcome and outcome complement configurations. For the latter, this suggests the unique ability of the evaluation function to generate disaggregated identity dimension outcome information *during* an evaluation is more important than having access to disaggregated information before an evaluation. This bears further research because it suggests that several mainstays included in the federal government's evaluation guidance on GBA+ may require updating

Future Gender-based Analysis Plus (GBA+) and evaluation research

In some respects, the impact of GBA+ within the Canadian federal government as a tool to advance EDI has never been more necessary and simultaneously, GBA+ falls significantly short of the haloed vision in achieving equity for all Canadians to date. Many of the newly introduced GBA+ policy requirements are now mandatory, but little systematic research has been undertaken to explore the process and product stemming from these integration efforts. Future research will advance knowledge on how the Canadian federal government plans for and uses GBA+ across its policy cycle as described above emphasizing value identification, production, and use focusing on the evaluation and performance measurement functions. Both the nascency and transparency of the Trudeau government's approach to GBA+ information provide a very

supportive research environment. There is little recognition of the availability of information and provides an innovative opportunity to examine how democracies prioritize EDI. A future program of study would propose to look at the following two projects

It has become clear in this research that only the performance measurement and evaluation functions adopt an ex-post or retrospective GBA+ lens. This means that only these two functions report retrospectively on *actual* gender and diversity results directly stemming from federal government expenditures at a variety of outcome levels (Whynot, 2018). The GBA+ annexes included in recent budget documents are selective in the inclusion of their retrospective and offer only limited analyses. The public availability of evaluations with the mandated inclusion of GBA+ stemming from the 2016 *Policy on Results* (TBS, 2016) provides a research-rich environment by which to explore GBA+ quality, the extent of integration, methods for including GBA+, and ultimately a more comprehensive understanding of how various evaluation functions are integrating GBA/GBA+ in a manner that supports instrumental evaluation use. As previously noted, evaluations must be made publicly available on the organizational website; it would therefore be simple enough to identify, retrieve, review, and analyze each of these evaluations to highlight exactly what instrumental use might be occurring because of specific items identified in Management Response Action Plans (MRAP). Discrete possible GBA+ and evaluation research projects might include the following:

- *Review of published federal government evaluation reports:* Beginning in 2009 as a result of Federal Accountability Act (FAA) legislation, the *Policy on Evaluation* (TBS, 2009) requires federal government organizations to publish final evaluation reports including MRAPs following organization approval processes. The identified period to publish evaluations is 120 days (TBS, 2009). This study would review publicly accessible

evaluation reports from all federal government organizations published since the implementation of the *Policy on Results* (TBS, 2016) using a GBA+ lens to determine how gender and diversity were considered. Planning for a comprehensive review of evaluation reports in a few years would also allow for the integration of GBA+ in an organization's 5-year Departmental Evaluation Plan (DEP) and would cover the grandfathering extension allowing evaluation functions to transition to new policy requirements. The study would seek to identify the following elements in the evaluation reports reviewed: GBA+ conceptual issues identified; the integration approach related to evaluation core issues; levels and types of intersectional research identified, target populations identified and served; and perhaps most importantly, the results achieved. This project would also provide a unique opportunity to examine federal government decision-makers' instrumental and deliberate non-use of produced GBA+ information, program amendments suggested, and progression towards related evaluation recommendations.

- *Use of GBA+ in Evaluation:* This project would undertake interviews with Heads of Evaluation (n=30). Given that sufficient time has passed since the articulation of GBA+ as a specific federal government priority in the *Policy on Results* (Treasury Board Secretariat, 2016), it is reasonable to assume that evaluation leaders will be able to identify the potential impacts of GBA+ on EDI on the evaluation function, but also further afield within various federal government organizations. Findings from this potential study will extend research regarding instrumental GBA+ and build on the value as it relates to conceptual, process, symbolic, and instrumental uses [Whynot, 2018].

Future Gender-based Analysis Plus (GBA+) and performance measurement research

The federal government's performance measurement function is also responsible for ex-post GBA+ data collection, that is, demonstrating exactly what government interventions have done regarding diversity and inclusion. As a consequence of the implementation of the *Policy on Results* (TBS, 2016), Heads of Performance Measurement (HoPM) positions were created within each federal government organization to address longstanding challenges regarding the lack of performance measurement information to support evaluation efforts (Cousins & Aubry, 2006; Treasury Board Secretariat, 2004; Treasury Board Secretariat, 2005). Guidance materials accompanying the release of the *Policy on Results* (TBS, 2016) suggest that organizations should integrate GBA+ in both their Departmental Results Frameworks (DRF), as well as their Program Information Profiles (PIPs). Three projects proposed within this stream respond to queries on the presence and extent of integration:

- *Survey of Heads of Performance Measurement*: This project would feature a survey of the HoPM across the federal government to assess the state of integration of GBA+ in performance measurement efforts. This survey would capture organizational knowledge/information about GBA+ including values, benefits, supports/enabling factors including tools and training, risks and mitigation strategies, capacities and use elements. Survey findings will inform subsequent projects in this stream.
- *Review of GBA+ resources in Performance Measurement*: This project encompasses a review of organizations GBA+ results planning and reporting efforts including DRFs, PIPs, and Reports on Plans and Priorities (RPP): Although not the focus of the candidate's doctoral research, unsurprisingly, a relationship was identified regarding the availability of GBA+ performance measurement information and the extent of GBA+

integration in evaluation. Organizations that intentionally included GBA+ in performance measurement efforts reflected stronger evaluation efforts related to GBA+. Relevant documents will be reviewed, and analysis undertaken to determine the extent to which GBA+ was integrated into these results planning and reporting efforts. Key research questions will focus on the following elements: the extent to which GBA+ was considered; target populations identified; metrics associated with GBA+ measurement; the value anticipated as well as realized, and identified level of effort. This project also represents uncharted research domains and could potentially inform not only policy development processes domestically, but internationally as well.

- *Use of GBA+ in Performance Measurement:* This project will undertake interviews with the Heads of Performance Measurement (n=30). Given that sufficient time has passed since the articulation of GBA+ as a specific federal government priority in the *Policy on Results*, it is reasonable to assume that performance measurement leaders will be able to identify the potential impacts of integration of GBA+ on the performance measurement function not only within their respective functional communities but more broadly within their organizations. The findings from this project will examine novel value formations, values identified and acted upon as well as the potential uses stemming integration of GBA+ in performance measurement.

Final Reflections

In conclusion, insights are offered on both GBA/GBA+ as a tool and its integration in the federal government evaluation function. Without progression in the first area, only minimal advancement will be possible in the latter. Historically, the process of prioritizing gender in policy discourse has in the past been broadly referred to as gender mainstreaming. This must

change with the introduction and advanced understanding of intersectionality research. It was not until mid-way through this research (WAGE, 2019) that related public-facing GBA+ materials de-emphasized gender as the primary analytical variable allowing this variance to align with organizational mandates. The addition of a simple “+” obviates all other identity dimensions experienced by individuals in their interactions with their socio-political, economic, and cultural environments. It fails to capture the complexity with all dimensions of an individual’s identity and how they engage with their environment which results in an experience that is unique to them. Failing to name identity elements contributes to a need for individuals with diverse personal identity elements to be recognized. It is consequently misleading, and disrespectful to suggest that all identity elements could coalesce under GBA/GBA+. Intersectionality methodologists (Scott & Siltanen, 2017) make the case that to preclude identity elements at the onset of investigative efforts is to eliminate potentially critical information that could be used to inform decision-making. Continuing to lead in this way, situates gender and/or sex as the most important analytical variables which is not always the case (Ragin & Fiss, 2017). Overall, an understanding of intersectionality, and intersectional research is lacking in the federal government and particularly in the evaluation function.

The success of the GBA/GBA+ tool lies in its stimulation of discussions across the bureaucracy in ways that have not been evidenced historically. However, its failure lies inherently in its lack of value articulation. Traditional approaches to decision-making within democratic societies are predicated on principles related to Western conceptualizations of Enlightenment, in which rationality assumes centre stage. However, early prominent evaluation theorists such as Weiss (1973, 1998, 1989) have long noted that evaluation use by decision-makers as one “piece of the political game” (1998, p.9) in which normative trade-offs are

required. The European Commission recently tabled a report *Understanding Our Political Nature: How to Put Knowledge and Reason at the Heart of Political Decision-Making* that identifies the value of identity and the emotions driving decision-making. It simultaneously notes that these dimensions are not properly understood nor debated (European Commission, 2019). Cumulatively, GBA+ information flowing from evaluation processes and products is merely one input into decision-making. This invalidates the belief that supplying sufficient information about differential outcomes will lead to more informed decision-making which in turn, will lead to improved equity outcomes. What is missing in the policy discourse discussion surrounding the attendant value of GBA/GBA+ to support EDI outcomes.

Although the Trudeau government's attention focus on GBA/GBA+ has reprioritized its longstanding commitment to gender equality, the "+" elements are emphasized to a lesser extent including within the federal government evaluation function. At the onset of this research, when GBA had not yet become GBA+, it was envisioned that one of the outputs would be a competency framework identifying the various aspects of integration and associated levels of competency. It was realized that undertaking such an effort would be problematic in light of the rapidly shifting dynamic of integrating GBA+ across the federal government policy cycle with the magnitude of changes realized, and impending. Many questions remain regarding the integration of GBA/GBA+ in the federal government evaluation cycle. Predominantly amongst these immediate questions is addressing how to facilitate reporting on contextual factors including socio-political, economic, and cultural aspects interacting with various identity dimensions which vary across borders both real and imagined for a national level program? If the federal government is serious about valuing various identity elements in their interventions, contextual factors will require greater consideration than they have been accorded to date. These

considerations may include the adoption of evaluation methods and approaches that extend beyond the federal government's current practice or even serve as an initiating call for the emergence of more innovative approaches to evaluation that centralize intersectionality theory and practice.

List of Appendices

Appendix A	Ethics clearance and renewal dates
Appendix B	Heads of Evaluation survey
Appendix C	Key informant interview guide
Appendix D	List of key informants and affiliation
Appendix E	Case study questions (pre-post)
Appendix F	Case study participant minutes
Appendix G	Methodological checklist for integrating GBA+

List of Appendices

Appendix A	Ethics clearance and renewal dates
Appendix B	Heads of Evaluation survey
Appendix C	Key informant interview guide
Appendix D	List of key informants and affiliation
Appendix E	Case study questions
Appendix F	Case study participant minutes
Appendix G	Methodological checklist for integrating GBA+

Appendix A: Ethics Information

Ethics File Number: 11-16-21

Renewed: January 9, 2019

Closed: February 26, 2020

Appendix B: General consent and introduction to all data collection

You are invited to participate in the research study “*Where’s Betty?: Integrating Gender-Based Analysis in the Canadian Federal Government Evaluation Function.*” There are six pages following this one. It will take approximately **20 minutes** to complete this survey.

The research study is conducted by Jane Whynot, Ph.D. Candidate, and supervised by Dr. Caroline Andrew at the University of Ottawa. If you have any questions about this research, you may contact the researcher or her supervisor at the coordinates provided below:

Lead Researcher: Jane Whynot, Ph.D. Candidate
Institute of Feminist and Gender-Based Analysis Studies, Faculty of Social Sciences,
University of Ottawa
120 University, Social Sciences Building, Room 11002
Ottawa, Ontario, K1N 6N5
Phone [REDACTED] (*obscured to meet University privacy policies*)
Email [REDACTED] (*obscured to meet University privacy policies*)

Supervisor: Dr. Caroline Andrew
Director Centre on Governance, Faculty of Social Sciences, University of Ottawa
120 University, Social Science Building, Room 5043
Ottawa, Ontario, K1N 6N5
Phone (613) 520-5800 x2755
Email [REDACTED] (*obscured to meet University privacy policies*)

Purpose of the Study: This research explores what integrating Gender-Based Analysis into the federal government’s evaluation function involves. Information will be collected through key informant interviews, a survey of individuals responsible for the evaluation function in individual departments, case studies, and a research review.

Key research questions include:

- How do stakeholders identify the relevance and value of integrating Gender-Based Analysis in the federal government’s evaluation function?
- What is required from various stakeholders throughout the evaluation process to integrate Gender-Based Analysis in the federal government’s evaluation function?
- Where in the evaluation lifecycle can Gender-Based Analysis be integrated into the federal government’s evaluation function?

- How does the inclusion of Gender-Based Analysis in the federal government evaluation function impact the usefulness of evaluation?

Benefits: The potential benefits of participating in this study include gaining a better understanding how Gender-Based Analysis may be integrated in various evaluation stages in the government context and how to address related challenges.

Anonymity, Confidentiality and Risks: All information disclosed during research data collection will remain strictly confidential and used for the purposes of this research. Responses to open-ended questions may be used verbatim in presentations and publications but neither you (nor your organization) will be identified. This information in its entirety will be available only to the researcher and her supervisor. Respondent confidentiality will be protected by presenting results only in aggregate form. This will limit the potential risks to individual participants. In order to minimize the risk of security breaches and to help ensure your confidentiality we recommend that you use standard safety measures such as signing out of your account, closing your browser and locking your screen or device when you are no longer using them / when you have completed the study.

Safeguarding: Electronic versions of data will be kept in a secure manner. Electronic files will be encrypted and password protected; all survey data will be securely stored in the supervisor's office at the University of Ottawa for the full conservation period of five years following survey completion.

Accessibility: If you would like to participate in this survey as an individual responsible for the evaluation function in a federal department but cannot participate in the online survey, please contact the lead researcher at the coordinates above to request an alternative format of the survey.

Ethics Contact: If you have any questions regarding the ethical conduct of this study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 154, Ottawa, ON K1N 6N5, Tel.: (613) 562-5387 or email ethics@uottawa.ca.

Copy of Research: If you would like to receive a summary of the research when it has been completed, you may include your coordinates at the end of the survey. Please keep a copy of this information for your records. Thank you for your time and participation.

Voluntary Participation: This survey focuses on awareness and understanding of key commitments, capacities, experiences, and results from integrating Gender-Based

Analysis and evaluation. Your participation in this online survey is limited to responding to closed and open-ended responses. Your participation in this [X] is voluntary, and you do not have to answer any questions that you do not want to answer. If you elect to withdraw from the [X] after commencing, your responses will be removed from the sample.

[specific paragraph for each data collection tool and associated consent requirement]

Regards,
Jane Whynot

Appendix C: Heads of Evaluation Survey

Where's Betty?
*Integrating Gender-Based Analysis (GBA) in the Canadian Federal
 Government Evaluation Function*

Gender-Based Analysis Awareness, Tools and Training

	Yes	No
Is your organization's evaluation function <u>conceptually aware</u> of gender-based analysis?	☐	☐
Is your organization's evaluation function <u>familiar with the principles</u> of gender-based analysis?	☐	☐
Is your organization's evaluation function <u>comfortable applying the principles</u> of gender-based analysis?	☐	☐
Has your organization's evaluation function received any <u>training</u> on gender-based analysis?	☐	☐
Is your organization aware of Status of Women Canada's Gender-Based Analysis Plus tool "GBA+"?	☐	☐
Has your organization's evaluation function taken the Status of Women Canada's GBA+ training?	☐	☐
Have gender policies, initiatives or related guidance materials specific to your organization been developed?	☐	☐
• Do you believe that your organization's evaluation function is affected by these policies, initiatives or related guidance materials?		
Is your organization's evaluation function aware of existing national commitments to integrate gender-based analysis in assessments of program, policies, legislations, and initiatives?	☐	☐

Gender-Based Analysis Perceptions: Relevance, Value, Benefit

	Yes	No
Is it relevant for your organization's evaluation function to include gender-based analysis in their evaluation practice?	☐	☐

Yes No

Can your organization's evaluation function identify the value of including gender-based analysis in their evaluation practice?

☐ ☐

Can your organization's evaluation function identify the benefits of including gender-based analysis in their evaluation practice?

☐ ☐

In your organization's evaluation function, is gender-based analysis understood to be "feminist"?

☐ ☐

- In your organization's evaluation function, does "feminist" carry a connotation?
- Is the connotation positive or negative?

Does your organization's evaluation function believe that the objectives of providing government decision-makers with information of both gender-based analysis and evaluation are aligned?

☐ ☐

Does your organization's evaluation function believe that the assessment functions associated with gender-based analysis and evaluation are aligned?

☐ ☐

Do you believe that the products (Budget proposals, Memoranda to Cabinet, Treasury Board Submissions) resulting from integrating GBA+ into evaluation align with policy cycle requirements?

Gender-Based Analysis and Evaluation Policy Supports

Yes No

Is your organization's evaluation function generally supportive of incorporating gender-based analysis in their evaluation practice?

☐ ☐

Has your organization's evaluation function noted the inclusion of gender-based analysis in the new federal government's *Policy on Results* (2016) accompanying *Directive on Results*?

☐ ☐

Has your organization's evaluation function noted the inclusion of gender-based analysis in other policy cycle requirements such as Budget proposals, Memoranda to Cabinet, Treasury Board Submissions?

☐ ☐

Yes No

Is your organization's evaluation function more inclined to incorporate gender-based analysis in evaluation if there are associated policy cycle requirements?

☐ ☐

Is your organization's senior management generally supportive of the evaluation function including gender-based analysis in their evaluation practice?

☐ ☐

Within your organization, do you believe it's reasonable given the current state of the evaluation function to incorporate gender-based analysis into evaluation practice?

☐ ☐

Tools to Integrate Gender-Based Analysis in Evaluation

Yes No

Does your organization's evaluation function consistently integrate gender-based analysis in their evaluation practice?

☐ ☐

Does your organization's evaluation function have the necessary financial resources to support integrating gender-based analysis in their evaluation practice?

☐ ☐

Does your organization's evaluation function have sufficient capacity to integrate gender-based analysis in their evaluation practice?

☐ ☐

Does your organization's evaluation function have access to internal expertise to support integrating gender-based analysis in their evaluation practice?

☐ ☐

Has your organization's evaluation function consulted their Gender Focal Point for assistance integrating gender-based analysis in their evaluation practice?

☐ ☐

Has your organization's evaluation function consulted their Gender Champion for assistance integrating gender-based analysis in their evaluation practice?

☐ ☐

Does your organization's evaluation function have access to external expertise to

☐ ☐

Yes No

support integrating gender-based analysis in their evaluation practice?

Where would your organization's evaluation function look, or with whom would you consult to obtain information on how to include gender-based analysis in evaluation?

Where would your organization's evaluation function look, or with whom would you consult to find federal government specific information on integrating gender-based analysis in evaluation?

Yes No

Has your organization developed tools to support integrating gender-based analysis in their evaluation practice?

☐ ☐

Does your organization's evaluation function have access to tools to support integrating gender-based analysis in their evaluation practice?

☐ ☐

Is your organization's evaluation function aware of the federal government's guidance on integrating Gender-Based Analysis Plus (GBA+) in evaluation?

☐ ☐

Does your organization's evaluation function have access to the federal government's guidance on integrating Gender-Based Analysis Plus (GBA+) in evaluation?

☐ ☐

Utilization Practice and Conditions

Yes No

Do your organization's senior decision-makers typically consider gender or other elements of diversity (such as age, geographic location, race, socio-economic class, religion) in their decision-making processes?

☐ ☐

Are you aware of any evaluations where your organization's senior decision-

☐ ☐

Yes No

makers have asked for information on gender or any other intersections of diversity?

Do your organization's program managers typically consider gender or other elements of diversity (such as age, geographic location, race, socio-economic class, religion) in their decision-making processes?

☐ ☐

Are you aware of any evaluations where your organization's program managers have asked for information on gender or any other intersections of diversity?

☐ ☐

Of the evaluations that your organization's evaluation function has undertaken, do any of the interventions specifically target an element of diversity?

☐ ☐

How would you describe your organization's evaluation function's capacity to integrate gender-based analysis in evaluation? (low, developing, intermediate, established)

Research has identified five continuum points for undertaking intersectional research. They include:

- Single axis (one identity variable)
- Comparative (one identity variable, limited categories)
- Additive (more than one identity variable, limited categories)
- Interactional (more than one identity variable, multiple categories)
- Intersectional (multiple identity, multiple categories)

	Single axis	Comparative	Additive	Interactional	Intersectional
Now	☐	☐	☐	☐	☐
In 2 years	☐	☐	☐	☐	☐

Yes No

Does your organization's evaluation function systematically collect disaggregated performance measurement information for multiple identity axes?

☐ ☐

- Does your organization consider this to be problematic?

Survey Conclusion

How would you classify your organization's size?

- ☐ Large Department or Agency
- ☐ Small Department or Agency
- ☐ Micro Department or Agency

Of the sectors broadly used to represent areas of federal government delivery, which one does your organization represent?

- ☐ Economic Affairs
- ☐ Social Affairs
- ☐ International Affairs
- ☐ Government Affairs

On average, how many evaluations does your organization's evaluation function undertake each year?

	Yes	No
Would you be interested in participating in a key informant interview?	☐	☐

Would you like a copy of the results of this research? ☐ Yes ☐ No

Are there other comments that you would like to make?

Thank you.

If you have any questions about this research, you may contact the researcher at the coordinates provided below:

Lead Researcher: Jane Whynot, Ph.D. Candidate

Institute of Feminist and Gender Studies, Faculty of Social Sciences, University of Ottawa

120 University, Social Sciences Building, Room 11002

Ottawa, Ontario, K1N 6N5

Phone (obscured to meet University privacy policies)

Email (obscured to meet University privacy policies)

Appendix D: Key Informant Interview Guide Questions

Awareness and Understanding of the Integration of Gender and Evaluation

1. Please describe your role with respect to gender and evaluation.
2. Please describe your understanding of gender, evaluation or the integration of gender in evaluation.
3. How familiar are you with gender-based analysis? How familiar are you with GBA+ (the tool used by the Canadian federal government?)
4. How would you describe the purpose of gender-based analysis?
5. How would you describe the purpose of integrating gender-based analysis in evaluation?
6. How would you describe the relevance of integrating gender-based analysis in evaluation?
7. Do you believe integrating gender-based analysis in evaluation is an appropriate forum (in the government context) to question assumptions regarding systemic barriers encountered by various program beneficiaries?
8. How would you address concerns about embedding a situated feminist knowledge into a neutral/objective function?
9. How would you describe the ideal relationship between gender-based analysis and evaluation?
10. How would you describe the difference between the ideal and what is currently in practice in your jurisdiction?
11. What do you believe are the benefits of integrating gender, or gender-based analysis in evaluation to various stakeholders?

Supports, Resistances and Mitigation Strategies

12. Do you believe that having an overarching evaluation policy that specifically identifies gender-based analysis changes evaluation practice? How so?
13. What do you believe to be the facilitating factors that enable the integration of gender-based analysis in evaluation?
14. What resources are required to support integrating gender-based analysis in evaluation?
15. How would you describe the minimal criteria for integrating gender, or gender-based analysis into evaluation?
16. What are the common resistances that have been encountered in integrating gender-based analysis in evaluation?
17. What strategies have been identified to address these resistances?

Management Environment and Use of Information Resulting from the Integration of Gender and Evaluation

18. Can you provide an example in your jurisdiction where gender, or gender-based analysis was successfully integrated into evaluation?
19. Do you believe that the inclusion of gender-based analysis increases evaluation's quality, reliability and validity?
20. Do you believe that the quality of this information increased management's confidence in their subsequent decision-making?
21. Do you believe that integrating gender-based analysis in evaluation is more efficient than undertaking separate gender-based analysis reports, or separate evaluations?

22. Do you believe that it has the potential to become more efficient as it is used more frequently within your jurisdiction?
23. How were the findings generated from this integration of gender-based analysis used?
24. Were there subsequent changes in processes resulting from the integration of gender, or gender-based analysis in evaluation?
25. Can you recall any examples in which findings related to the integration of gender, gender-based analysis were linked to policy discussions, decisions or changes in policy?

Appendix E: List of key informants, affiliation and line of evidence

Name	Organization Affiliation	Position	Line of Evidence
Colleen Goggin	Atlantic Canada Opportunities Agency	Evaluation Manager	Key Informant - Practitioner
Ian Rankin	Canadian Institutes of Health Research	Head of Evaluation	Case Study
Sarah Viehbeck	Canadian Institutes of Health Research	Head of Performance Measurement	Case Study
Cara Tannenbaum	Canadian Institutes of Health Research/McGill University	Director, Institute of Sex and Gender	Case Study
Janet Siltanen	Carleton University	Professor Emeritus	Key Informant – Academic
Stephanie Paterson	Concordia University	Assistant Professor	Key Informant Academic - GBA
Adriana Newbury	Department of Canadian Heritage	Evaluation Project Officer	Case Study
Amanda Shannon	Department of Canadian Heritage	Head of Evaluation	Case Study
Julie Seguin	Department of Canadian Heritage	GBA Focal Point	Case Study
Luc Martin	Department of Canadian Heritage	Program Manager	Case Study
Catherine Bennett	Economic and Social Development Canada	Head of Performance Measurement, GBA+ Lead	Case Study
Christine Minas	Economic and Social Development Canada	Director of Evaluation	Case Study
Donna Smith-Montcrieffe, and Vincent Paquet	Economic and Social Development Canada	Evaluation Manager, Senior Evaluation Officer	Case Study
Yves Gingras	Economic and Social Development Canada	Head of Evaluation	Case Study
Jennifer Greene	University of Illinois	Past President American Evaluation Association	Key Informant – Academic
Svetlana Negroustoueva	EvalGender+	Co-Chair	Key Informant – Practitioner
Andrea Cyr	Fisheries and Oceans Canada	Head of Evaluation	Case Study
Ian Allan	Fisheries and Oceans Canada	GBA Working Group	Case Study
Stacey Prieur	Fisheries and Oceans	Senior Evaluation	Case Study

and Kathryn Radford	Canada	Officers	
Susan Cole	Fisheries and Oceans Canada	Director Results and Impact	Case Study
David Kurhors and Umit Kitilzan	Immigration, Refugees and Citizenship Canada	Head of Performance Measurement/Head of Evaluation	Key Informant – Practitioner
Alexandra Dagger and Ghislaine Tremblay	National Research Council of Canada	Chief Audit/Eval Executive, Head of Evaluation	Key Informant – Practitioner
Alfred LeBlanc	Natural Resources and Engineering Council Of Canada - Social Science and Humanities Research Council	Head of Performance Measurement	Case Study
Danika Goosney	Natural Resources and Engineering Council of Canada - Social Science and Humanities Research Council	GBA Champion/Vice President Research	Case Study
Jean-Francois Fortin	Natural Resources and Engineering Council of Canada - Social Science and Humanities Research Council	Senior Policy Analyst	Case Study
Serge Villemure	Natural Resources and Engineering Council of Canada - Social Science and Humanities Research Council	Director	Case Study
Susan Morris	Natural Resources and Engineering Council of Canada - Social Science and Humanities Research Council	Head of Evaluation	Case Study
Stephanie Kalt	Natural Resources Canada	Evaluation Manager, GBA+	Key Informant – Practitioner
Michael Bamberger	Rockefeller Foundation	Consultant/Author - Development Bank Advisor, Guidance	Key Informant – Practitioner
Daniel Sansfacon	Status of Women Canada	Former Head of Evaluation	Key Informant – Academic/ Practitioner
Linda Savoie	Status of Women Canada	Former Senior Director General	Key Informant – Practitioner

		Program, Head of Evaluation	
Marcelle Chabot	Status of Women Canada	Manager GBA	Key Informant – Practitioner
Suzanne Clement	Status of Women Canada	Former Head of Agency	Key Informant – Practitioner
Sebastian Goupil	Status of Women Canada, United Nations	Former Director General Policy, GBA Lead	Key Informant – Practitioner
Chris Boughton	Treasury Board Secretariat	Director of Results Management Group	Key Informant – Evaluation
Donna Podems	University of Michigan, and University of Johannesburg	Assistant Professor	Key Informant – Academic

Appendix F: Case Study Questions

Pre-effort Case Study Questions	Post-effort Case Study Questions
General Introduction 1. What are you hoping that your organization will gain from participating in this research? 2. What is your organization's general understanding about integrating gender-based analysis and evaluation?	1. What did your organization gain from participating in this research? 2. Did your organization's understanding about integrating gender-based analysis in evaluation change over the course of this research? How so?
Organizational Context 3. Why is your project team considering implementing gender-based analysis in this evaluation? 4. Does it make a difference to your project team that gender-based analysis is specifically referenced as an example in the new suite of policy tools accompanying the <i>Policy on Results (2016)</i>? 5. Do you think that the example referenced in the <i>Policy on Results (2016)</i> is sufficient to influence organizations' decisions to include gender-based analysis in evaluation? Please explain. 6. Does the project team conceptually understand how gender-based analysis can be integrated in this evaluation? 7. Is the project team comfortable in applying the principles of gender-based analysis in this evaluation? 8. Has the project team had gender-based analysis training? 9. Do you understand how gender-based analysis can be integrated in evaluation practice empirically? 10. Do you believe that the purposes of the gender-based analysis and evaluation functions are aligned?	3. How were the project team's considerations in flagging this evaluation for integrating gender-based analysis in evaluation met? 4. How has the project team's conceptual understanding of integrating gender-based analysis in evaluation increased as a result of their experiences? How so? 5. How have the project team's comfort levels in applying the principles of gender-based analysis increased as a result of their experience? How so? 6. To what extent did project team participate in any gender-based analysis training? 7. To what extent did the project team's empirical understanding of integrating gender-based analysis in evaluation change? How so? 8. To what extent did your project team's understanding about the alignment of purposes between gender-based analysis and evaluation change? Please explain how. 9. To what extent did your project team's understanding about the assessment dimensions associated with integrating gender based analysis and evaluation

11. Do you believe that the assessment dimensions associated with integrating gender based analysis and evaluation functions are aligned? 12. Do you believe that the information provided by gender-based analysis and evaluation has the potential to support management's decision-making processes? Please elaborate. 13. Are you aware of your organization having an organization wide policy or process addressing gender-based analysis? Please explain. 14. Do these tools affect/have the potential to affect your organization's evaluation function? How so? Please explain.	functions change? Please explain how. 10. To what extent was the information provided by integrating gender-based analysis in evaluation useful to organization decision-makers? 11. To what extent did your project team's understanding about the value of information provided by gender-based analysis and evaluation to support senior management's decision-making processes change? Please explain how. 12. To what extent has your organization implemented an organization wide policy or process addressing gender-based analysis since beginning this research? Please explain. 13. To what extent did these tools affect the organization's evaluation function in any way? How so? Please explain. 14. Are you now aware of other evaluations in your organization having experience undertaking gender-based analysis in their evaluation efforts?
Organization Awareness and Support 15. Do you believe that stakeholders (program management, senior decision makers, evaluators) are generally aware of federal commitments to undertake gender-based analysis? How so? Please explain. 16. Do you believe that organization stakeholders (senior decision-makers and evaluators) are aware of efforts to integrate gender-based analysis in evaluation? 17. Are organization stakeholders (senior decision-makers and evaluators) supportive of efforts to integrate gender-based analysis into evaluation? Why or why not? 18. How do organization stakeholders (senior	15. To what extent do you believe that stakeholders (program management, senior decision makers, evaluators) generally awareness of federal commitments to undertake gender-based analysis has changed? How so? Please explain. 16. To what extent do you believe that organization stakeholders (senior decision-makers and evaluators) awareness of efforts to integrate gender-based analysis into evaluation has changed? How so? Please explain. 17. How have organization stakeholders (senior decision-makers and evaluators) support of efforts to integrate gender-

decision-makers and evaluators) demonstrate this support? Please explain.	based analysis into evaluation changed? Why or why not? 18. How has organization stakeholders' (senior decision-makers and evaluators) demonstration of support changed? Please explain.
<i>Resources and Capacities</i> 19. Does your evaluation unit have experience integrating gender-based analysis that you are aware of? Please describe this experience. 20. Does your project team have experience integrating gender-based analysis in evaluation that you are aware of? 21. Have you set aside or allocated related resources for this project? What are these resources? 22. What expertise do you intend to access for this specific effort?	19. To what extent does your organization's evaluation unit now have access to gender-based analysis capacities that wasn't available prior to this research? What are these capacities? Please explain. 20. To what extent does your organization's evaluation unit now have access to gender-based analysis expertise that wasn't available prior to this research? What is this expertise? Please explain. 21. To what extent do you think that your evaluation unit has sufficient experience integrating gender-based analysis? Please explain.. 22. What other related resources could your organization's evaluation unit use? 23. To what extent were related resources created either during, or after this research's conclusion? 24. To what extent were evaluation resources set aside to support integrating gender-based analysis sufficient for this initiative? 25. To what extent was the gender-based analysis expertise that you had intended to access for this evaluation engaged in this effort? Did you engage other gender expertise that wasn't anticipated being engaged? 26. Was this expertise internal or external to the organization?
<i>Opportunities and Challenges</i> 23. What are the supportive mechanisms in	27. What were the most supportive

place that lead you to believe that this program is an appropriate candidate to integrate gender-based analysis into evaluation? Please explain. 24. What do you anticipate the challenges to be for this effort? Please explain. 25. Is disaggregated program performance information available? Can disaggregated program performance information be collected throughout the evaluation data collection phase? Please explain.	mechanisms relied on during this initiative? Did these supports align with your original identification? Please explain. 28. What were the challenges associated with this initiative? Please explain. 29. Did new challenges emerge that were not anticipated? Please explain what these new challenges were. 30. To what extent was having/not having access to disaggregated program performance information a challenge? Was this concern addressed during the evaluation? How was this concern addressed?
<i>Stakeholder Engagement</i> 26. What organizational stakeholders do you anticipate engaging in this effort to integrate gender-based analysis and evaluation? Are these different from the stakeholders that you normally engage with during an evaluation? How are they different? What is the anticipated value of their engagement? 27. Do stakeholders anticipate making special accommodations in the evaluation planning phase to integrate gender-based analysis in evaluation? What are the accommodations that you anticipate making in this phase? 28. Do stakeholders anticipate making special accommodations into the evaluation data collection phase? What are the accommodations that you anticipate making in this phase? 29. Do stakeholders anticipate making special accommodations into the evaluation analysis phase? What are the accommodations that you anticipate making in this phase?	31. What organizational stakeholders were engaged to integrate gender-based analysis and evaluation? What was the value of their engagement? 32. To what extent did stakeholders make special accommodations in the evaluation planning phase to integrate gender-based analysis in evaluation? What were these accommodations? 33. To what extent did stakeholders make special accommodations in the evaluation data collection phase? What were these accommodations? 34. To what extent did stakeholders make special accommodations in the evaluation analysis phase? What were these accommodations? 35. To what extent did stakeholders make special accommodations in the evaluation reporting phase? What were these accommodations? 36. To what extent was the value of these accommodations realized? Please explain the actual value of any accommodations

30. Do stakeholders anticipate making special accommodations into the evaluation reporting phase? What are the accommodations that you anticipate making in this phase? 31. What is the anticipated value of these accommodations? 32. Which of these stakeholders do you believe will be influential in this specific effort? Please explain.	to the initiative. 37. Which of these stakeholders was influential to this specific initiative? Why were they influential?
<i>Utility</i> 33. How important is integrating gender into evaluation for this specific project? Why? 34. What are the intended benefits of integrated gender-based analysis into this project? 35. Do you foresee any other ways by which to include gender in evaluation for this project?	38. To what extent were the benefits of integrating gender-based analysis into this effort realized? Please explain. Were other benefits realized? 39. To what extent did decision-makers use/do they intend to use information resulting from integrating gender-and evaluation that you are aware of? How is this information being used?

Appendix F: Case Study Participant Minutes

Summary Discussion Points from Friday June 8, 2018: Integrating Gender-based Analysis *Plus* (GBA+) Emerging Promising Practices

Agenda

- Organizations asked to share experiences across the evaluation planning, conducting, analysis and reporting phases that included the identification of both enabling, and inhibiting factors at each of these evaluation phases. Some of the organizations have not yet moved beyond the planning phase while others have implemented at the individual evaluation level.
- Scope of the case studies identified that included two organizations that elected to build GBA+/Eval capacity in various ways, and four organizations that proceeded on a single specific evaluation case basis.
- A few individuals from organizations not included in the case studies but either identified as innovators (at the leading edge of implementation) and just beginning planning for incorporating GBA+ across the organizations/in evaluation were invited to attend.

<i>Planning Phase</i>	• Various ways exist to approach including GBA+ in evaluation. Options range from systematically integrating GBA+ in, and across organizational evaluations, to integration efforts at the individual evaluation level. • At the individual evaluation level GBA+ can be undertaken in evaluations at various levels that include a) to those provided with funding to deliver program activities, and b) to those who are benefitting from the activities delivered by funders as well as c) organizationally. All are valid and depending on context, should be considered depending on what decision-makers needs are. • Integrating GBA+ in evaluations can be both direct and indirect. The former includes a “straight-up” questioning of gender equality, gender-based analysis etc. Indirectly, GBA+ information is produced/used through disaggregating information at the outcome/indicator data. Both have to be planned for. Often, organizations that perceive themselves as having less experience prefer to approach in an indirect manner. • GBA+ may be included in both relevance and performance dimensions of evaluation. Including GBA+ in design/delivery aspects of an evaluation drives to the needs of participants.
<i>Enabling Factors</i>	• Supports are critical to the enabling environment as it’s not just evaluation alone that requires moving GBA+ forward but needs to involve stakeholders from other functional areas including program, performance measurement, policy, strategic planning, risk etc. The supports can come in various forms such as physical (setting aside resources for each evaluation, disaggregated data), financial (foregone opportunity costs associated with time set aside to build related capacities), and symbolic with real-time implications (setting aside time for group discussion, leadership articulation and support for GBA+). • Specific supports referenced include training to build capacity, the identification, and internalization of gender equality (and GBA+) as a priority,

	leadership supports at various levels, amendment of organizational infrastructure (to capture disaggregated data), supporting stakeholder engagement across the organization and engaging those that may not have directly engaged with one another prior to these related opportunities. • The policy supports (like GBA+ input into the Budget, Memorandum to Cabinet (MCs), Treasury Board Submissions (TB Subs) as well as mandate letters have also been instrumental in moving GBA+ forward, and more specifically in integrating GBA+ in evaluation. This is now a requirement beyond the evaluation function. • To support disaggregated data collection, some organizations in their planning efforts have negotiated for third party organizations (those funded) to collect GBA+ information through funding agreements in place with the federal/provincial/territorial committees. This requires pre-planning to identify and obtain necessary information. • Plans are also in the works to use GBA+ information culled from Departmental Results Frameworks (DRFs) and Program Information Profiles (PIPs) to inform forthcoming evaluation efforts. • It also requires relationship building with those responsible. This is also relevant to those administering Gs&Cs but a tension exists between collecting GBA+ information and streamlining reporting requirements. • No matter the extent of the supports, timing must also be suitable to initiate these efforts. Oftentimes, a confluence of factors have contributed to making the timing right including both federal government and organizational priorities. • The more information available on GBA+ elements related to the program, and its beneficiaries, the more helpful it is. Having this information can assist in scoping what element of intersectionality should be investigated in further detail. This may include literature reviews and reviews of the grey literature. • Both front-end and continuing efforts need to include building culture of leadership related to GBA+ that includes both educational and informational aspects: ○ Some leadership cadres are unaware of the potential of GBA+ and what is required to undertake this. They need to be provided with information (and examples of evaluations incorporating GBA+ which are in short supply) ○ Others are fully debriefed and have requested GBA+ specific information ○ Having performance management agreements for executives that have identified GBA+ commitments is also deemed as beneficial. • Several of the organizations have in place Equity, Diversity & Inclusion initiatives (EDI). Aligning GBA+ efforts with EDI efforts was believed to be advantageous. EDI documentation may be available and includes such resources as related organizational strategic plans and logic models. EDI designated groups are established and include (gender, visible minority, ability, and Aboriginal statuses). These elements align with GBA+ but are not exhaustive.
--	--

	• Several of the organizations have invested significant resources in planning to integrate GBA+ in evaluation but have not yet done so.
<i>Challenges</i>	• A main challenge regards the naming of the GBA+ tool/process itself. Gender and sex are designed to serve as the <i>entry</i> points to analysis with other intersectional dimensions of identity following, however, there is an overwhelming emphasis on gender and the other intersections of diversity fall to the side as a result. As well, some of the organizations have found it difficult to identify what dimension of the intersectional factors apply to them. There aren't sufficient resources to look at all GBA+ information and therefore efforts need to be investigated in advance to narrow this down. Scoping GBA+ is much like scoping an evaluation given finite resources for the function. • One of the challenges deals with privacy concerns. Self-identification is voluntary; participants do not have to disclose any personal information which can make it problematic to undertake rigorous, reliable GBA+. • It was identified that additional guidance materials both relating to GBA+ frameworks, and GBA+/evaluation would be beneficial. It was viewed as problematic by some that the <i>process</i> of developing frameworks isn't explored in more detail.
<i>Conducting Phase</i>	• GBA+ includes methods that may be more appropriate to some groups of stakeholders than others. GBA+ information should include both qualitative and quantitative information. GBA+ (the tool) forwarded by SWC suggests that they prefer the relational intersectional model of how personal identity factors relate which is associated with the validation of quantitative models. • It is important to be upfront in any relevant statements about why this data is being collected and how it is going to be used. This includes positioning GBA+ as a tool to address existing systemic barriers (and by extension – discrimination experienced historically by various members of diverse groups of women and men).
<i>Enabling factors</i>	• Where GBA+ information has been gathered at the front end of the application process those responsible for IT and privacy within individual organizations have been engaged in the process/discussions. Requesting GBA+ information as part of the funding opportunities applications may be problematic as it is self-disclosed data and not yet verified/verifiable. • There was believed to be a positive relationship between the organizations that included relevant indicators in their Departmental Results Frameworks (DRFs) and Program Information Profiles (PIPs) although this has not yet been realized. What it does indicate is management's attention and support to GBA+. • Evaluators are bright, analytical people and adding additional variables to the analysis component of evaluations has been supported by the function itself, and by others. The incorporation of GBA+ in evaluation competencies is thought to be well-aligned.
<i>Challenges</i>	• On surveys: Addressing privacy concerns, organizations have addressed various ways that include: at the front end – collecting personally identifiable information as part of the funding application process, and at the evaluation end - adding a preface regarding the disclosure of this information in their data collection tools. However, it may be the case that even when the wording of

	this preface is “right” – data may not result. This was the experience of some of the participating organizations. • There is some concern that asking participants to self-disclose personal identity information may cause perverse program effects.
Reporting Phase	• Ultimately, GBA+ is about reporting on differential impacts on diverse groups of women and men. SWC has directed that they are looking for a “number”. • There are some questions about how to present GBA+ information in the context of an evaluation report that include framing it separately, or integrating related findings. Some have presented it as a separate section within the report but embedded it within the findings. Others have planned for a cross-cutting GBA+ organizational study, however concerns exist regarding the comparison of GBA+ elements (like barriers). • Guidance materials (international and draft domestic) suggest balancing the presentation of GBA+ findings relevant to other findings, but ensuring the relevant recommendations are presented as specifically addressing GBA+. • As part of creating a culture of leadership, there needs to be communications directed to senior decision-makers on both the strengths and deficits of various elements of integrating GBA+ information (including disaggregated information). This should be based on evidence/experiences. This reflective element could potentially be culled from recommendations introduced in reporting elements stemming from the integration of GBA+ in evaluation. • The language that is used is important. Conceptually it’s important to differentiate between equity/equality (where equity has different identified needs as a starting point).
Enabling factors	• There is an opportunity to include the need regarding the lack of/ability to include disaggregated GBA+ relevant information in the recommendations as this can be later included in TB Subs and MCs to make the case for funding to support implementation • There is an existing current “hook” for senior decision-makers given federal government priorities in reporting at evaluation, and the various policy cycles to date that were previously discussed. This represents an opportunity to embed and demonstrate equality/equity through reporting by the various organizations. However, evaluation is not the only producer of GBA+ information at this juncture, and may never be the sole producer of GBA+ information.
Challenges	• In the future, it will be much easier to include GBA+ in reviews of program documentation including TB Subs and MCs given that programs will have been designed with a GBA+ element in mind. Currently, the evaluations planned for and occurring are dealing with older documents that did not refer to GBA+, nor was it required. Consequently, this could be viewed as a transitional phase.

Appendix G: Methodological Checklist for Integrating GBA+ in Evaluation

This tool is comprised of a series of questions that can be used at both the inception, and conclusion of each evaluation to help guide evaluation scoping/design questions and as a “check-in” to identify whether potential implementation strengths and weaknesses may exist. The results from the application of this tool may improve the integration of GBA+ in each evaluation, as well as improve the capacity of various evaluation functions when considered in conjunction with the GBA+ conceptual tools.

At the Evaluation Scoping/Pre-Design Phase

Methodological/Practical GBA+ Considerations	Yes	No	N/A	Explanation
Are there articulated GBA+ organizational priorities (GBA+ Action Plans or Frameworks) that should be considered in evaluation efforts?				
Is GBA+ information available from organizational performance measurement information that can inform the evaluation (Departmental Results Framework, Program Information Profile)?				
Is GBA+ information included in either the Memoranda to Cabinet (MC) or Treasury Board Submission (TB Sub)? What are these GBA+ elements?				
Is GBA+ information included in the evaluability assessment? What are these GBA+ elements?				
What are other jurisdictions doing with respect to serving various diverse groups?				
Is reach to diverse populations considered in the program logic model or theory of change?				

At the Evaluation Design Phase

Methodological/Practical Consideration	Yes	No	N/A	Explanation
All GBA+ items are clearly labelled in the evaluation matrix.				
Pertinent GBA+ elements have been examined and have been identified in the evaluation matrix.				
The evaluation questions and/or sub-questions present opportunities to explore GBA+ issues.				
The evaluation matrix contains indicators pertaining to GBA+ that are likely to provide information to support undertaking GBA+.				
The evaluation matrix reflects where the program is at in terms of GBA+ implementation (core dimensions reference from training materials).				
Decision makers' information needs regarding GBA+ are				

incorporated into the evaluation matrix.				
The evaluation methodology addresses identified GBA+ differences and provides opportunities for differences to be identified.				
The evaluation methodology provides opportunities for expression of diverse experiences and perspectives.				
The GBA+ information to be produced by the evaluation is both qualitative and quantitative.				
The GBA+ information to be produced by the evaluation is likely to be useful for reducing differences.				
Opportunities have been created for diverse populations to contribute to the evaluation processes on a voluntary basis (e.g., informing the evaluation design, providing evaluation information through participation in evaluation methods, participating in data collection, contributing to the analysis and validation of findings, reviewing the evaluation report, and/or providing feedback on the evaluation recommendations).				
The evaluation function has the resources (time, capacity, expertise, and finances) necessary to carry out the integration of GBA+, as detailed in the evaluation matrix.				

At the Evaluation Data Collection and Analysis Phases

Methodological/Practical Consideration	Yes	No	N/A	Explanation
The lines of evidence (e.g., document reviews, literature reviews, interviews, focus groups) include an examination of GBA+ elements, as well as intersectional elements.				
Alternative sources of information (e.g., alternative news sites, social media, conference proceedings) are searched to identify GBA+ differences				
The sample sizes for evaluation methods involving diverse populations e.g., for interview, focus group, survey participants) are representative of diverse populations or weighting has been considered				
Safeguards have been put in place to address concerns that may leave participants from diverse populations vulnerable in disclosing sensitive data (e.g., participants are informed of how the information they share will be used and protected).				
Only information that is necessary for the purpose of the evaluation is requested, to respect participants' privacy.				
The language employed to collect identity information takes intersectionality into account and is specific enough				

to obtain information at the level of granularity that is necessary for the purpose of the evaluation				
The data collection methods appropriately reflect the cultural and contextual factors of each participant diverse population (e.g., culturally appropriate use of open-ended and closed-ended interview questions, scheduling of focus group aligned with participants' availabilities).				
The language and content of the data collection tools are accessible to and respectful of all participant populations.				
The quantitative and qualitative analyses take into consideration each participant population's unique context, as well as the relationships between program stakeholder groups (internal and external) and the power dynamics that exist between these groups.				
The evaluation's GBA+ findings are validated by representatives of the diverse populations				

At the Evaluation Reporting Phase

Methodological/Practical Consideration	Yes	No	N/A	Explanation
The evaluation report includes a description of the methods used to integrate GBA+ into the evaluation approach and methodology.				
The ways in which representatives of diverse populations contributed to the evaluation process are described in the evaluation report.				
Evaluation resources dedicated to integrating GBA+ are described in the evaluation report.				
The added value of including GBA+ is described in the evaluation report.				
The evaluation findings address the needs of diverse populations, barriers to their program access and participation, programs resources and tools to remove barriers, and differences in program outcomes/impacts across diverse populations.				
GBA+ findings are presented in a balanced manner with fair representation of different perspectives.				
GBA+ findings are reflected in the evaluation's conclusions and recommendations.				
Findings and recommendations include the identification of data gaps and needs for disaggregated performance measurement data about GBA+ elements.				
The recommendations stemming from GBA+ findings identify means of reducing the inequities found between diverse populations				

The recommendations stemming from GBA+ findings reflect an evidence-based understanding of the unique contexts and cultures of diverse populations				
The Management Response and Action Plan adequately addresses the recommendations stemming from GBA+ findings.				
Strategies are implemented to disseminate the evaluation report among diverse populations.				
Lessons learned regarding integrating GBA+ into evaluation are documented by the AEB evaluation function				
Amendments required and made to AEB GBA+ and Evaluation Framework Tool and Supplementary Tool				

References

- Ahmadzai, M. (2015). *A Study on Visible Minority Immigrant Women's Experiences with Domestic Violence*. Masters Thesis Social Justice and Community Engagement Wilfred Laurier University.
- Akins, J. (2017). *Because its 2015 and we still don't understand GBA+: The problem of implementation for gender policy in the Government of Canada*. Masters Thesis Concordia University.
- Alkin M. & Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1–12. [http://doi: 10.1016/S0191-491X\(03\)90001-0](http://doi:10.1016/S0191-491X(03)90001-0).
- Amartya, S. (2004), Capability and well-being. In Nussbaum, M. & Sen, A. (eds.), *The quality of life* (pp.30-53). Routledge.
- Anderson, E. & Langford, M. (2012). *A distorted metric: The MDGs and state capacity*. Madrid: Third Annual Meeting of Human Rights Metrics.
- Babbie, E. & Benaquisto, L. (2002). *Fundamentals of Social Research*. 1st Canadian Edition, Thomson Canada Limited.
- Babbie, E. (1995). *The Practice of Social Research* (7th Ed.). Wadsworth Publishing Company.
- Bamberger, M., Mabry, L., & Rugh, J. (2020). *Realworld evaluation: Working under budget, time, data, and political constraints*. Sage Publications Inc..
- Bamberger, M., Segone, M. and Reddy, S. (2014). *National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems*.

https://www.evalpartners.org/sites/default/files/documents/evalgender/NationalEvaluationPolicies_web-single-color%281%29.pdf.

Barnes, J. & Bishop, J. (2016). *Review of UN SWAP Evaluation Performance Indicator Reporting Synthesis Report Final Version*.

<http://www.uneval.org/document/download/2634>.

Bemelmans-Videc, M., Rist, R. & Vedung, E. (1998). *Carrots, sticks & sermons: Policy instruments and their evaluation*. Transaction Publishers.

Bickman, Leonard. (1987). The functions of program theory. *New Directions for Program Evaluation*, 33, pp. 5–18. [https:// doi:10.1002/ev.1443](https://doi.org/10.1002/ev.1443).

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation*, 13(4), 439–455.
<https://doi.org/10.1177/1356389007082129>.

Borys, S., Gauthier, B., Kishchuk, N. & Roy, S. (2019). *Bridging 2015 & 2018: How has the Policy on Results modified federal evaluation practice?* Canadian Evaluation Society Annual Conference, Nova Scotia.

Bougie, E., & Kohen, D. (2018). Smoking correlates among Inuit men and women in Inuit Nunangat. *Health Reports*, 29(3), 3–10.

Bourgeois, I. & Whynot, J. (2018). The influence of evaluation recommendations on instrumental and conceptual uses: An exploratory study. *Evaluation and Program Planning*, 68, 13-18.

- Bourgeois, I., & Whynot, J. (2018). Strategic evaluation utilization in the Canadian federal government., *Canadian Journal of Program Evaluation*, 32(3), 327-346.
- Bretschneider, P., Cirilli, S., Jones, T., Lynch, S., & Wilson, N. Document review as a qualitative research data collection method for teacher research. *Sage Research Methods Cases*. <https://doi.10.4135/9781473957435>
- Brousselle, A., Buregeya, J. (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, 24(2), 153–168. <https://doi:10.1177/1356389018765487>.
- Bustelo, M. (2016). Evaluation from a gender + perspective as a key element in (re) gendering the policy-making process. *Journal of Women, Politics & Policy*. 38(1), 84-101.
- Canada, Parliament. (2017). *Interim status report on the implementation of the departmental action plan on gender-based analysis*. <https://cfc-swc.gc.ca/gba-accs/progress-etape-en.html>.
- Canada, Parliament. Standing Committee on Public Accounts. (2010). *Chapter 1; Gender-based analysis," of the spring 2009 report of the Auditor General of Canada..* 40th Parl. 3rd sess. https://www.ourcommons.ca/Content/Committee/403/PACP/Reports/RP4414397/403_PACP_Rpt04/403_PACP_Rpt04-e.pdf
- Canada, Parliament. Standing Committee on Public Accounts. (2010). *Governmental response to Chapter 1; Gender-based analysis," of the spring 2009 report of the Auditor General of*

Canada.. 40th Parl.3rd sess. <https://www.ourcommons.ca/DocumentViewer/en/40-3/PACP/report-4/response-8512-403-27>

Canada, Parliament. Standing Committee on the Status of Women. (2016). *Implementing gender-based analysis plus in the Government of Canada*. 42nd Parl., 1st sess. <https://www.ourcommons.ca/DocumentViewer/en/42-1/FEWO/report-4/>

Canada, Parliament. Standing Committee on the Status of Women. (2016). *Government Response to implementing gender-based analysis plus in the Government of Canada*. 42nd Parl., 1st sess. <https://www.ourcommons.ca/DocumentViewer/en/42-1/FEWO/report-4/response-8512-421-76>

Canada, Parliament. Standing Committee on the Status of Women. (2005). *Gender-based analysis: Building blocks for success*. 39th Parl., 1st sess. <https://www.ourcommons.ca/DocumentViewer/en/39-1/FEWO/report-2/>

Canada, Parliament. Standing Committee on the Status of Women. (2005/6) *Government response to gender-based analysis: Building blocks for success*. 39th Parl., 1st sess. <https://www.ourcommons.ca/DocumentViewer/en/38-1/FEWO/report-2/response-8512-381-112>

Canadian Charter of Rights and Freedoms, s 7, Part I of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (UK), 1982, c11.

- Carastathis, A. (2016). Interlocking systems of oppression. In Rodriguez, N., Martino, W., Ingrey, J. & Brockenbrough, E. (eds). *Critical concepts in queer studies and education: A guide for the twenty-first century*. Palgrave McMillan.
- Carvalho, S. and White, H. (1997). Combining the quantitative and qualitative approaches to poverty measurement and analysis: The practice and the potential. World Bank Technical Paper 366. Washington, D.C..
- Centre for Theories of Change.(2019). Setting Standards for Theories of Change.
<https://www.theoryofchange.org/what-is-theory-of-change/>
- Centre of Excellence for Evaluation Treasury Board Secretariat. (2004). Evaluation Function in the Government of Canada. <https://www.tbs-sct.gc.ca/cee/pubs/func-fonc-eng.asp>.
- Chalifour, N. (2017). How a gendered understanding of climate change can help shape Canadian climate policy. In *Climate Change and Gender in Rich Countries* (1st ed., pp. 233–249). Routledge. <https://doi.org/10.4324/9781315407906-15>
- Chant, S., & Sweetman, C. (2012). Fixing women or fixing the world `Smart economics`, efficiency approaches, and gender equality in development. *Gender & Development*, 20(3): 517-529.
- Chen, H. and P.H. Rossi (1980). Multi-goal theory-driven approach: A model linking basic and applied social science. *Social Forces*, 51(1), 106-122.
- Chen, H.-T., & Rossi, P. H. (1983). Evaluating With sense: The theory-driven approach. *Evaluation Review*, 7(3), 283–302. <https://doi.org/10.1177/0193841X8300700301>.

- Cheng, T., Nosova, E., Small, W., Hogg, R., Hayashi, K., & DeBeck, K. (2019). A gender-based analysis of nonmedical prescription opioid use among people who use illicit drugs. *Addictive Behaviors*, 97, 42–48. <https://doi.org/10.1016/j.addbeh.2019.05.022>.
- Clark, H. & Anderson, A. (2004). Logic models and theories of change: Telling them apart. American Evaluation Association Annual Conference. Atlanta, Georgia.
- Coen, S., Subedi, R., & Rosenberg, M. (2016). Working out across Canada: Is there a gender gap?: Working out across Canada. *The Canadian Geographer*, 60(1), 69–81. <https://doi.org/10.1111/cag.12255>.
- Collins, P. (2020). Intersectionality (2nd edition). Wiley Publishing.
- Conklin, A., Yao, C., & Richardson, C. (2019). Chronic sleep disturbance, not chronic sleep deprivation, is associated with self-rated health in adolescents. *Preventive Medicine*, 124, 11–16. <https://doi.org/10.1016/j.ypmed.2019.04.014>.
- Connell, J. & Kubisch, A. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. In K. Fulbright-Anderson, A. C. Kubisch, & J. P. Connell (Eds.), *New approaches to evaluating community initiatives: Volume 2. Theory, measurement, and analysis* (pp. 1-13). Washington, D.C.
- Cooke, M., Waite, N., Cook, K., Milne, E., Chang, F., McCarthy, L., & Sproule, B. (2017). Incorporating sex, gender and vulnerable populations in a large multisite health research

- programme: The Ontario Pharmacy Evidence Network as a case study. *Health Research Policy and Systems*, 15(1), 20–. <https://doi.org/10.1186/s12961-017-0182-z>.
- Council of Europe. (2004). Gender Mainstreaming: *Conceptual framework, methodology and presentation of good practices*. Strasbourg, France. <https://rm.coe.int/1680596135>.
- Cousins, B. & Bourgeois, I. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*. 34, 299-319.
- Crenshaw, Kimberlé W. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, Feminist Theory and Antiracist Politics, 1989 University of Chicago Legal Forum, pp. 139–67.
- Creswell, J. & Poth, C. (2018). *Qualitative Inquiry & Research Design: Choosing among five approaches*. Sage Publications Inc.
- Darnton, A. (2008). *Government Social Research (GSR) behaviour change knowledge review: practical guide: An overview of behaviour change models and their uses*.
<https://research.fit.edu/media/site-specific/researchfitedu/coast-climate-adaptation-library/climate-communications/messaging-climate-change/Darnton.-2008.-Behaviour-Change-Models--Uses.pdf>
- Davies, Rick. (2012). *Criteria for assessing the evaluability of theories of change*.
mandenews.blogspot.com/2012/04/criteria-for-assessing-evaluability-of.html.
- Denzin, Norman K. (1973). *The research act: A theoretical introduction to sociological methods*. Transaction Publishers.

DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd ed.). Sage Publications Inc.

Devers, K., Lallemand, N., Burton, R., Kahwati, L., McCall, N., & Zuckerman, S. (2013). Using qualitative comparative analysis (QCA) to study patient-centered medical homes. U.S. Centers for Medicare & Medicaid Services.

<https://www.urban.org/sites/default/files/publication/24261/412969-Using-Qualitative-Comparative-Analysis-QCA-to-Study-Patient-Centered-Medical-Homes.PDF>.

Dhillon, L. & Vaca, S. (2018). Refining theories of change. *Journal of MultiDisciplinary Evaluation*, 14(30), 64-87. https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/496>.

Dillman, D., Smyth, J. & Christian, L (2014). *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.). John Wiley & Sons Inc.

Doerr, A. (2006). *Royal Commission on Aboriginal Peoples*. Toronto, ON: Historica Foundation. DOI 10.7208/Chicago/9780226414546.001.0001

Drabble, J. & McInnes, S. (2017). Finding her home a gender-based analysis of the homelessness crisis in Winnipeg. https://www.policyalternatives.ca/sites/default/files/uploads/publications/Manitoba%20Office/2017/03/Finding_Her_Home_%20low-res.pdf.

Eichler, M. (1987). *Sexism in Research and its Policy Implications: No. 6 Policy Brief*. Canadian Institute for Advancement of Women.

- Eichler, M. (1997). Feminist methodology. *Current Sociology*, 45(2), 9–36. <https://doi.org/10.1177/001139297045002003>
- Espinosa, J., Bustelo, M., & Velasco, M. (2016). Evaluating gender structural change: Guidelines for evaluating gender equality action plans. GENOVATE/Complutense University of Madrid. http://www.genovate.eu/media/genovate/docs/intconferencebrusselsdocs/GENOVATE_Guidelines_for_evaluating_GEAPs.pdf.
- Fetterman, D. (1994). Empowerment evaluation. *Evaluation Practice*, 15(1), 1–15. [https://doi.org/10.1016/0886-1633\(94\)90055-8](https://doi.org/10.1016/0886-1633(94)90055-8).
- Findlay, T. (2019). Revisiting GBA/GBA+: Innovations and interventions. *Canadian Public Administration*, 62(3), 520–525.. <https://doi.org/10.1111/capa.12341>.
- Fleischer, D. & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association Members. *American Journal of Evaluation*, 30(2), 158–175. <https://doi.org/10.1177/1098214008331009>.
- Forss, K., Rebien, C., & Carlsson, J. (2002). Process use of evaluations: Types of use that precede lessons learned and feedback. *Evaluation*, 8(1):29-45.
- Fukuda-Parr, S., Yamin, A. & Greenstein, J. (2014). The power of numbers: A critical review of Millennium Development Goal targets for human development and human rights. *Journal of Human Development and Capabilities*, 15(2-3), 105-117.
- Funnell, S., & Rogers, P. (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey-Bass.

- Galport, N., & Azzam, T. (2016). Evaluator training needs and competencies. *American Journal of Evaluation*, 38 (1), 80–100. <https://doi.org/10.1177/1098214016643183>.
- Government of Canada. (2013-2018). Capacity survey on results functions. <https://open.canada.ca/data/en/dataset/681f4124-0dbd-400a-8d0c-3fa21c3b1775>
- Grant, M. & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi:10.1111/j.1471-1842.2009.00848.x>.
- Greaves, L. (2020). Missing in Action: Sex and Gender in Substance Use Research. *International Journal of Environmental Research and Public Health*, 17(7), 2352. <https://doi.org/10.3390/ijerph17072352>.
- Greene, J. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, 12(2), 91–116. <https://doi.org/10.1177/0193841X8801200201>.
- Greene, J. (2007). *Mixed Methods in Social Inquiry*. Jossey-Bass.
- Haarmans, M., McKenzie, K., Kidd, S., & Bentall, R. (2018). Gender role strain, core schemas, and psychotic experiences in ethnically diverse women: A role for sex- and gender-based analysis in psychosis research? *Clinical Psychology and Psychotherapy*, 25(6), 774–784. <https://doi.org/10.1002/cpp.2307>.
- Hankivsky O. *Health inequities in Canada: intersectional frameworks and practices*. Vancouver, BC: UBC Press; 2011.

Hankivsky, O. & Cormier, R. (2010). Intersectionality and public policy: Some lessons from existing models. *Political Research Quarterly*, 64(1), 217-229.

<https://doi.org/10.1177/1065912910376385>.

Hankivsky, O. & Mussell, L. (2018). Gender-Based Analysis Plus in Canada: Problems and Possibilities of Integrating Intersectionality. *Canadian Public Policy*, 44(4), 303–316.

<https://doi.org/10.3138/cpp.2017-058>.

Hankivsky, O. (2007). Gender Mainstreaming in the Canadian Context: One Step Forward and Two Steps Back, in M. Orsini and M. Smith (eds.). *Critical Policy Studies*. University of British Columbia Press.

Hankivsky, O. (2012). *An intersectionality-based policy analysis framework*. Institute for Intersectionality Research and Policy, Simon Fraser University.

Hankivsky, O. (2012). The lexicon of mainstreaming equality: Gender-based analysis (GBA), gender and diversity analysis (GDA) and intersectionality-based analysis. *Canadian Political Science Review*, 6(2/3), 171-183.

Hankivsky, O. (2013). Gender mainstreaming: A five-country examination. *Politics and Policy*, 41(5), 629-655.

Harding, A. (2018). The effect of government transfer programs on low-income rates: A gender-based analysis, 1995 to 2016, income research paper series. Statistics Canada.

<https://www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2018003-eng.htm>.

- Head, B. (2018). Understanding Wicked Policy Problems. *Policy Options*. <https://policyoptions.irpp.org/magazines/january-2018/understanding-wicked-policy-problems/>.
- Health Communication Capacity Collaborative. SBCC and Gender Theories. sbccimplementationkits.org/gender/sbcc-gender-theories/.
- Hoang-Kim, A., Parpia, C., Freitas, C., Austin, P., Ross, H., Wijesundera, H., Tu, K., Mak, S., Farkouh, M., Sun, L., Schull, M., Mason, R., Lee, D., & Rochon, P. (2020). Readmission rates following heart failure: a scoping review of sex and gender based considerations. *BMC Cardiovascular Disorders*, 20(1), 223–19. <https://doi.org/10.1186/s12872-020-01422-3>
- Hood, D. & Cassaro, D. (2002). Feminist evaluation and the inclusion of difference. *New Directions in Evaluation*, 96(3), 3-8. <https://doi.10.1002/ev.64>
- Hopson, R. (1999). Minority issues in evaluation revisited: Re-conceptualizing and creating opportunities for institutional change. *American Journal of Evaluation*, 20, 445-451. <https://doi:10.1177/109821409902000304>.
- House, E. (1999). Evaluation and people of colour: A response to Professor Stanfield. *American Journal of Evaluation*, 20, 433-435. DOI: 10.1177/109821409902000302
- I., Donaldson Stewart. (2007). *Program theory-driven evaluation science: Strategies and applications*. Taylor & Francis.

- Johnson, K., Greenseid, L., Toal, S., King, J., Lawrenz, F., & Volkov, B. (2009). Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410. <https://doi.org/10.1177/1098214009341660>.
- Johnstone, R. & Momani, B. (2019). Organizational change in Canadian public institutions: The implementations of GBA+ in DND/CAF. *Canadian Public Administration*, 62(3), 500–519. <https://doi:10.1111/capa.12338>.
- Jost, C., Kristjanson, P. & Ferdous, N. (2015). Lessons in theory of change: Gender and inclusion CCSL Learning Brief No.14. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). https://www.pacificclimatechange.net/sites/default/files/documents/CGIAR_Theory%20of%20change_GenderInclusion.pdf
- Kabeer, N. (1999). *United Nations Research Institute for Social Development Discussion Paper 108: The Conditions and Consequences of Choice – Reflections on the Measurement of Women`s Empowerment*. United Nations Research Institute for Social Development.
- Kahwati, L. & Lane, H.(2020). *Qualitative Comparative Analysis in mixed methods research and evaluation*. California Sage Publications.
- Kennedy, M., Hayashi, K., Milloy, M., Boyd, J., Wood, E., & Kerr, T. (2020). Supervised injection facility use and exposure to violence among a cohort of people who inject drugs: A gender-based analysis. *The International Journal of Drug Policy*, 78, 102692–. <https://doi.org/10.1016/j.drugpo.2020.102692>.

- Kirkpatrick, D. L. (1959). Techniques for evaluation training programs. *Journal of the American Society of Training Directors*, 13, 21-26.
- Kloosterman, J., Benning, E. & Fyles, R. (2012). Measuring the unmeasurable: Gender mainstreaming and cultural change. *Gender & Development*, 20(3): 531-545.
- Knott, J. & Wildavsky, A. (1980). If dissemination is the solution, what is the problem? *Knowledge*, 1(4), 537–578. <https://doi.org/10.1177/107554708000100404>.
- Koleros, A. & Mayne, J. (2019). Using actor-based theories of change to conduct robust contribution analysis in complex settings. *Canadian Journal of Program Evaluation*, 33(3), 292-315. <https://doi.org/10.3138/cjpe.52946>.
- Koleros, A., Mulkerne, S., Oldenbeuvig, M. & Stein, D. (2018). The actor-based change framework: A pragmatic approach to developing program theory for interventions in complex systems. *American Journal of Evaluation*, 41(1), 34–53. <https://doi.org/10.1177/1098214018786462>.
- Kvale, S. and Brinkmann, S. (2015). Interviews: Learning the craft of qualitative research interviewing (3rd edition). Sage Publications.
- Lacasse, A., Pagé, M., Choinière, M., Dorais, M., Vissandjée, B., Nguefack, H., Katz, J., Samb, O., & Vanasse, A. (2020). Conducting gender-based analysis of existing databases when self-reported gender data are unavailable: the GENDER Index in a working population. *Canadian Journal of Public Health*, 111(2), 155–168. <https://doi.org/10.17269/s41997-019-00277-2>

Laing, G. (2013). *Royal Commission on bilingualism and biculturalism*. Toronto, ON: Historical Foundation.

Lam S, Dodd W, Whynot J, Skinner K. (2019). How is gender being addressed in the international development evaluation literature? A meta-evaluation. *Research Evaluation*. 28(2): 158-168.

Langford, M. (2012). *The art of the impossible: Measurement choices and the post-2015 development agenda background paper*. New York: United Nations Development Programme.

Lanthier, S., Bielecky, A., & Smith, P. (2018). Examining Risk of Workplace Violence in Canada: A Sex/Gender-Based Analysis. *Annals of Work Exposures and Health*, 62(8), 1012–1020. <https://doi.org/10.1093/annweh/wxy066>.

Leviton, L.C., & Hughes, E.F.X. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review*, 5, 525-549.

Lewis, J. & Grzanka, P. (2017). *Intersectional theories*. In K. Nadal (Ed.) *The SAGE encyclopedia of psychology and gender* (pp. 969-971). SAGE Publications, Inc. doi: 10.4135/9781483384269.n325

Lockwood, T. (1967). *The history of Royal Commissions*. *Osgoode Hall Law Journal*, 5(2): 172-209.

- Lombardo, E. & Mergaert, L. (2013). Gender mainstreaming and resistance to gender training: A framework for studying implementation. *Nordic Journal of Feminist and Gender Research*, 21(4), 296-311. <https://doi.10.1080/08038740.2013.851115>.
- Lorde, A. (1979). The master's tools will never dismantle the master's house. Personal and the Political Panel of New York University Institute for the Humanities Conference. New York: New York.
- MacDonnell, V. (2018). Gender-based analysis plus (gba+) as constitutional implementation. *Canadian Bar Review*, 96(2).
- Mackenzie, M. and A. Blamey (2005). The practice and the theory: Lessons from the application of a theories of change approach. *Evaluation*, 11(2), 151—68.
- Maillé, C. (2018). Intersectionalizing gender policies: experiences in Quebec and Canada. *French Politics*, 16(3), 312–327. <https://doi.org/10.1057/s41253-018-0058-9>
- Mark, M. (1990). From program theory to tests of program theory. *New Directions for Program Evaluation*, 47, 37–51. <https://doi:10.1002/ev.1553>.
- Marra, M. (2003). *Dynamics of evaluation use as organizational knowledge: The case of the World Bank*. Dissertation Abstracts International: Section A: The Humanities and Social Sciences, 64, 1070 (UMI 3085545).
- Mayne, John. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation*, 30(2), 119–142. <https://doi:10.3138/cjpe.230>.

Mayne, John. (2017). The COM-B Theory of Change Model (V3).

McCall, L. (2001). *Complex inequality: Gender, class, and race in the new economy*. Routledge..

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society*, 30(3), 1771–1800. <https://doi:10.1086/426800>.

McCarthy, L., Milne, E., Waite, N., Cooke, M., Cook, K., Chang, F., & Sproule, B. (2017). Sex and gender-based analysis in pharmacy practice research: A scoping review. *Research in Social and Administrative Pharmacy*, 13(6), 1045–1054. <https://doi.org/10.1016/j.sapharm.2016.11.007>

McNutt, K. & Béland, D. (2015). Implementing an integrated governance strategy: The quest for gender mainstreaming in Canada. *American Review of Canadian Studies*, 45(4), 467-485. <https://doi:10.1080/02722011.2015.1116591>.

McNutt, K. (2010). *An integrated approach to gender equality: From gender-based analysis to gender mainstreaming working paper series issue 2*. Johnson-Shoyama Graduate School of Public Policy.

Meier, P. & Lombardo, E. (2013). Gender quotas, gender mainstreaming and gender relations in politics. *Political Science*, 65(1), 46–62. <https://doi:10.1177/0032318713488114>.

Mertens, D. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation*, 20(1), 1-14. <https://doi:10.1177/109821409902000102>

- Mertens, D. (2014). A transformative feminist stance: Inclusion of multiple dimensions of diversity with gender. In Brisolara, S., Seigart, D., & SenGupta, S. (Ed.) *Feminist evaluation and research: Theory and practice* (pp.95-112). The Guildford Press.
- Mertens, D. M., Farley, J., Madison, A. M., & Singleton, P. (1994). Diverse voices in evaluation practice: Feminists, minorities, and persons with disabilities. *Evaluation Practice*, 15(2), 123-129.
- Michie, S., vanStralen, M. & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(42). <https://doi.org/10.1186/1748-5908-6-42>.
- Montague, S. & Porteous, N. (2013). The case for including reach as a key element of program theory. *Evaluation and Program Planning*, 36(1), 177–183. <https://doi:10.1016/j.evalprogplan.2012.03.005>.
- Morris, C. (2006). *Royal Commission on the status of women in Canada*. Toronto, ON: Historical Foundation.
- Morris, M., Clark, B. (2013). You want me to do what? Evaluators and the pressure to misrepresent findings. *American Journal of Evaluation*, 34, 57–70.
- Moser, C. (1989). Gender planning in Third World: Meeting practical and strategic gender needs. *World Development*, 17(11): 1799-1825.
- Müller-Clemm and Barnes. (1997). A historical perspective on federal program evaluation in Canada, *Canadian Journal of Program Evaluation*, 12(1), 47-70.

Natcher, D., Maria Bogdan, A., Lieverse, A., & Spiers, K. (2020). Gender and Arctic climate change science in Canada. *Palgrave Communications*, 6(1), 1–8.

<https://doi.org/10.1057/s41599-020-0407-6>

National Capital Chapter Canadian Evaluation Society. (2019). Annual Learning Event Program.

<https://ncc.evaluationcanada.ca/wp-content/uploads/2019/02/ALEprogram2019-Full.pdf>

Nussbaum, M. (2011). *Creating capabilities*. Cambridge, MA: Harvard University Press.

Office of the Auditor General. (2009). Spring Report of the Auditor General of Canada: Chapter One Gender-Based Analysis. Ottawa: Auditor General.

Office of the Auditor General. (2015). Fall Report of the Auditor General of Canada: Chapter One Gender-Based Analysis. Ottawa: Auditor General.

Parliament of Canada. (2005). *Equality for women beyond the illusion: The expert panel on accountability mechanisms for gender equality final report*. Ottawa: Canada.

Parliament of Canada. (n.d.). Mandate of the Standing Committee on the Status of Women.

Accessed August 2016: <http://www.parl.gc.ca/Committees/en/FEWO/About> .

Parliament of Canada. *Report on the Royal Commission on the status of women in Canada*. (1970). Ottawa: Canada.

Paterson, S. (2010). What's the problem with gender-based analysis? Gender mainstreaming policy and practice in Canada. *Canadian Public Administration*, 53(3), 395– 416.

Patton, M.Q. (2008). *Utilization-focused evaluation* (4th edition). Sage Publications Inc.

Peck, L. R. & Gorzalski, L. M. (2009). An evaluation use framework and empirical assessment.

Journal of Multidisciplinary Evaluation. 6(12): 139-156.

Perezneito, P. & Taylor, G. (2014). A review of approaches and methods to measure economic empowerment of women and girls. *Gender & Development*, 22(2): 233-251.

Planning and Performance Exchange 2019 Symposium. (2019). <https://ppx.ca/en/symposiums-eng-2019/>.

Podems, D. (2010). Feminist evaluation and gender approaches: There's a difference? *Journal of MultiDisciplinary Evaluation*, 6(14), 1-17.

Podems, D. (2014). Feminist evaluation for nonfeminists. In Brisolara, S., Seigart, D., & SenGupta, S. (Eds.) *Feminist Evaluation and Research: Theory and Practice* (pp.113-142). New York: The Guildford Press.

Prime Minister's Office of Canada. (2017). Minister of Status of Women Mandate Letter.

Prime Minister's Office of Canada. (2017). President of Treasury Board Secretariat Mandate Letter.

Privy Council Office of Canada. (2018). Mandate letter tracker: Delivering results for Canadians. Accessed June 2018: <https://www.canada.ca/en/privy-council/campaigns/mandate-tracker-results-canadians.html>

Privy Council Office, The Treasury Board Of Canada Secretariat & Status Of Women Canada. (2009). *Departmental action plan on gender-based analysis in response to audit findings*

and recommendations contained in chapter 1, “Gender-based Analysis” of the spring 2009 report of the auditor general of Canada. <https://cfc-swc.gc.ca/gba-accs/ap-pa-eng.pdf>.

Public Health Agency of Canada. (2019). 2019–20 Departmental Plan: Public Health Agency of Canada. <https://www.canada.ca/en/public-health/corporate/transparency/corporate-management-reporting/reports-plans-priorities/2019-2020-report-plans-priorities.html>.

Ragin, C. & Fiss, P. (2017). *Intersectional inequality: Race, class, test scores and poverty*. University of Chicago Press.

Raynor, M. (1986). Using evaluation in the federal government. *Canadian Journal of Program Evaluation*, 1, 1-10.

Rees, T. (1998). *Mainstreaming Equality in the European Union*. Routledge/

Rich, R. F. (1977). Measuring knowledge utilization: Processes and outcomes. *Knowledge and Policy*, 10, 11–24.

Rihoux, B. & Meur, G. (2009). Crisp-set qualitative comparative analysis (csQCA). In Rihoux, B., & Ragin, C. (eds.) *Applied Social Research Methods: Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques* (Vol. 51, pp. 33-68). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781452226569

Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and related systematic comparative methods: Recent advances and remaining challenges for social science research. *International Sociology*, 21, 679–706.

- Ritzer, G. (2007). *Contemporary sociological theory and its classical roots: The basics*. Boston: McGraw-Hill.
- Robeyns, I. (2016). Capabilitarianism. *Journal of Human Development and Capabilities*, 17(3), 397-414. <https://doi:10.1080/19452829.2016.1145631>.
- Rochette, A. (2016). Climate change is a social justice issue: the need for a gender-based analysis of mitigation and adaptation policies in Canada and Quebec. *Journal of Environmental Law and Practice*, 29, 383–410.
- Rogers, P. (2000). Causal models in program theory evaluation. *New Directions for Evaluation*, 87, 47–55. <https://doi:10.1002/ev.1181>.
- Rowlands, R. (1997). *Questioning empowerment: Working with women in Honduras*. Oxfam Great Britain.
- Rutman, L. (1986). Some thoughts on federal level evaluation. *Canadian Journal of Program Evaluation*, 1, 19-27.
- Scala, F. & Paterson, S. (2017). Bureaucratic role perceptions and gender policy work in Canada. *Gender, Work and Organization*, 24(6), 579-593.
- Scheaffer, R., Ott, L, & Mendenhall, W. (1986). *Elementary survey sampling* (3rd ed). PWS-Kent, Boston.
- Schneider, C. Q., Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to Qualitative Comparative Analysis (QCA)*. Cambridge University Press.

Scott, N. & Siltanen, J. (2012). Gender and intersectionality – A quantitative toolkit for analyzing complex inequalities. Employment and Social Development Canada.

Scott, N. & Siltanen. (2016). Intersectionality and quantitative methods: Assessing regression from a feminist perspective. *International Journal of Social Research Methodology*, 20(4), 373–385. <https://doi:10.1080/13645579.2016.1201328>.

Scriven, M. (1991). *Evaluation Thesaurus*. Sage.

Scriven, M. (2011). Evaluation, bias and its control. *Journal of MultiDisciplinary Evaluation*, 7(15), 79-98.

Segsworth, R. (2005). *Program evaluation in the Government Of Canada: Plus ca change. Canadian Journal of Program Evaluation*. 20(3), 175–197.

Shadish, W. (1987). Program micro- and macrotheories: A guide for social change. *New Directions for Program Evaluation*, 33, 93–109. <https://doi:10.1002/ev.1448>.

Shulha, L., & Cousins, B. (1997). Evaluation use: theory, research, and practice since 1986. *Evaluation Practice*. 18(3),195–208. [https://doi: 10.1016/S0886-1633\(97\)90027-1](https://doi: 10.1016/S0886-1633(97)90027-1).

Sielbeck-Bowen, K., Brisolara, S., Seigart, D., Tischler, C. & Whitmore, E. (2002). Exploring feminist evaluation: The ground from which we rise. *New Directions in Evaluation*, 96(3), 3-8. <https://DOI: 10.1002/ev.62>.

Silverman, D. (2020). *Interpreting qualitative research: A guide to the principles of qualitative research* (5th edition). Sage Publications Inc.

Skinner, K. (2007). Developing a tool to measure knowledge exchange outcomes. *Canadian Journal of Program Evaluation*, 22(1), 49-73.

Stake, R. (2005). *Qualitative case studies. The Sage handbook of qualitative research*. Sage Publications Inc.

Statistics Canada. (2018). The Centre for Gender, Diversity and Inclusion Consultation. <https://www.statcan.gc.ca/eng/consultation/cgdis>.

Status of Women Canada GBA+ online training Module 4: On the Ground. https://www.swc-cfc.gc.ca/gba-acsc/course-cours-2017/eng/mod04/mod04_09_01.html.

Status of Women Canada. (1998). Gender-based analysis: A guide for policy-making.

Status of Women Canada. (2005). An evaluation of three components of the Agenda for Gender Equality within Status of Women Canada: Final report by CS/Reesors.

Status of Women Canada. (2016). Action Plan on Gender-based Analysis 2016-2020.

Status of Women Canada. (2016). GBA+ annual implementation survey results 2016-17.

Status of Women Canada. (2017). Building a framework to support GBA+ in your organization: Policy statement.

Status of Women Canada. (2017). GBA+ research guide: Value of GBA+ research.

Status of Women Canada. (2017). Making GBA+ sustainable.

- Status of Women Canada. (2018). Evaluation of the Women's Program and Management Response. <https://cfc-swc.gc.ca/trans/account-resp/pr/wpeval-evalpf/index-en.html>
- Status of Women Canada. (2018). What is Gender-based Analysis Plus (GBA+) Module 3. https://cfc-swc.gc.ca/gba-acsc/course-cours/eng/mod03/mod03_02_01.html
- Status of Women Canada.(2018). *Evaluation of the Women's Program*. (2018). <https://cfc-swc.gc.ca/trans/account-resp/pr/wpeval-evalpf/index-en.html>.
- Stephens, A., Lewis, E. & Reddy, S. (2018). Towards an inclusive systemic evaluation for the SDGs: Gender equality, environments and marginalized voices (GEMs). *Evaluation*, 24(2), 220-236. [https://DOI: 10.1177/1356389018766093](https://doi.org/10.1177/1356389018766093).
- Stufflebeam, D. L. (1967). The use and abuse of evaluation in Title III. *Theory into Practice*, 6(3), 126-133.
- Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation theory, models, and applications*. Jossey-Bass.
- Suchman, E. (1971). Evaluating Educational Programs In Caro, Francis G. (Ed.), *Readings in Evaluation Research*. New York: Russell Sage Foundation, Pp. 43-48.
- Sutherland, N., Ward-Griffin, C., McWilliam, C., & Stajduhar, K. (2016). Gendered Processes in Hospice Palliative Home Care for Seniors With Cancer and Their Family Caregivers. *Qualitative Health Research*, 26(7), 907–920. <https://doi.org/10.1177/1049732315609571>.

Sutherland, N., Ward-Griffin, C., McWilliam, C., & Stajduhar, K. (2017). Structural impact on gendered expectations and exemptions for family caregivers in hospice palliative home care. *Nursing Inquiry*, 24(1), e12157–n/a. <https://doi.org/10.1111/nin.12157/>

Swedish International Development Cooperation Agency. (2002). *SIDA Studies No.3: Discussing women`s empowerment – Theory and practice*. Sweden.

The Expert Panel on Accountability Mechanisms for Gender Equality. (1995). Equality for Women: Beyond the Illusion: The expert panel on accountability mechanisms for gender equality final report. ISBN 0-662-42490-5

Thiem, A. (2017). Conducting configurational comparative research with Qualitative Comparative Analysis: A hands-on tutorial for applied evaluation scholars and practitioners. *American Journal of Evaluation*, 38(3), 420– 433. <https://doi.org/10.1177/1098214016673902>.

Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative Inquiry*, 17(6), 511–521. <https://doi.org/10.1177/1077800411409884>.

Treasury Board Secretariat of Canada. (2019). *Integrating Gender-Based Analysis Plus (GBA+) in Evaluation: A Primer*. <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/gba-primer.html#H-04>.

Treasury Board Secretariat. (1981). *Guide on the Program Evaluation Function*. <https://www.tbs-sct.gc.ca/cee/pubs/guide1981-eng.asp>

Treasury Board Secretariat. (1983). *Principles for the Evaluation of Programs by Federal Departments and Agencies*. <https://www.tbs-sct.gc.ca/cee/pubs/principles-principes1984-eng.asp>.

Treasury Board Secretariat. (2004). *Report on Effective Evaluation Practices* <https://www.tbs-sct.gc.ca/cee/tools-outils/eep-pee-eng.pdf>.

Treasury Board Secretariat. (2004). *Review of the Quality of Evaluations Across Departments and Agencies*. <https://www.tbs-sct.gc.ca/cee/pubs/review-examen2004-eng.asp#ftn2>.

Treasury Board Secretariat. (2009). *Policy on Evaluation*. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024>

Treasury Board Secretariat. (2010). *2010 Annual Report on the Health of the Evaluation Function*. <https://www.tbs-sct.gc.ca/report/orp/2011/arhef-raefetb-eng.asp>

Treasury Board Secretariat. (2011). *2011 Annual Report on the Health of the Evaluation Function*. <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/2011-annual-report-health-evaluation-function.html>

Treasury Board Secretariat. (2012). *2012 Annual Report on the Health of the Evaluation Function*. <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/2012-annual-report-health-evaluation-function.html>

Treasury Board Secretariat. (2015). Results-based Management Lexicon. <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/results-based-management-lexicon.html>.

Treasury Board Secretariat. (2016). *Policy on Results*. <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300>

Treasury Board Secretariat. (2017). *Interim Guidance on the Policy on Results*.

Treasury Board Secretariat. Evaluation Function in the Government of Canada. <https://www.tbs-sct.gc.ca/cee/pubs/func-fonc-eng.asp#:~:text=The%20goal%20of%20the%20Evaluation,produce%20better%20results%20for%20Canadians>.

Tuhiwai-Smith, L. (2012). *Decolonizing methodologies: Research and Indigenous Peoples*. New York: Palgrave McMillan.

United Nations Children's Fund. (2011). *How to design and manage equity-focused evaluations*.

United Nations Children's Fund. (2006). *The state of the world's children: Women and children, the double dividend of gender equality*.

United Nations Evaluation Group (UNEG). (2011). *Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance*. <http://www.uneval.org/document/download/1294>.

United Nations Evaluation Group (UNEG). (2014). *Integrating Human Rights and Gender Equality in Evaluations*. <http://www.uneval.org/document/download/2107>.

United Nations Evaluation Group (UNEG). (2016). *Norms and Standards for Evaluation*. <http://www.uneval.org/document/download/2613>.

United Nations Evaluation Group. (2018). UN-SWAP Evaluation Performance Indicator Technical Note Guidance Note. <http://www.uneval.org/document/download/2148#:~:text=UN%2D%20SWAP%201.0%20had%2015,and%20knowledge%20exchange%20and%20networking>.

United Nations Women and the Independent Evaluation Office. (2015). *How to manage gender-responsive evaluation: Evaluation handbook*.

United Nations Women Asia-Pacific. (2014). *Challenges of integrating gender equality in evaluation*. Malaysia.

United Nations Women. (1995). *Beijing Declaration and Platform for Action: Fourth World Conference on Women*.

Van eerdewijk, A. Brouwers, J. for HIVOS. (2015). Gender and theories of change. https://www.researchgate.net/publication/282976319_Gender_and_theories_of_change.

Verloo, M. (2001). Another velvet revolution? Gender mainstreaming and the politics of implementation. IWM Working Paper No. 5/2001. Vienna.

- Weiss, C. (2016). Have we learned anything new about the use of evaluation? *The American Journal of Evaluation*, 19(1), 21–33. <https://doi.org/10.1177/109821409801900103>.
- Whynot, J. (2018). *Bridging the divide: The potential relationship between GBA+ use and its quality*. American Evaluation Association 2018 Conference. Cleveland, Ohio.
- Whynot, J. (2018). Bridging the divide: The relationship between gender-based analysis "plus" (GBA+) use and its quality. European Evaluation Society 13th Biennial Conference. Thessaloniki, Greece.
- Whynot, J. (2018). Emerging promising practices on integrating gender-based analysis in evaluation. Planning and Performance Exchange (PPX) Annual Symposium. Ottawa.
- Whynot, J., Lemire, S. & Montague, S. (2019). How we model matters: A manifesto for the next generation of program theorizing. *Canadian Journal of Program Evaluation*, 33(3), 414-4433. <https://doi:10.3138/cjpe.53070>.
- Whynot, J., Mavriplis, C., Farenhorst, A., Langelier, E., Franz-Odendaal, T. & Shannon, L. (2019). Knitting theory in STEM: Experiences in developing a performance framework. *Canadian Journal of Program Evaluation*. 33(3): 354-374.
- World Bank. (1994). *Enhancing women`s participation in economic development, Report No.: 12972 GLB*. World Bank.