Evaluation Capacity Case Study: Organizational Assessment and Recommendations to Build Evaluation Capacity

By

Leah Johnson
B.A., Political Studies, University of Alberta, 2013

A Master's Project Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS IN COMMUNITY DEVELOPMENT

in the School of Public Administration

Defense Committee

Client:	Sharon Yeo, Director Immigration and Settlement Service, Catholic Social Services
Supervisor:	Dr. Astrid Brousselle School of Public Administration, University of Victoria
Second Reader:	Dr. Jill Anne Chouinard School of Public Administration, University of Victoria
Chair:	Dr. Tara Ney School of Public Administration, University of Victoria

Acknowledgements

First, I want to thank Dr. Dittmar Mundel for being an exceptional teacher and mentor in my undergraduate degree at Augustana Faculty, University of Alberta. His life and work as an educator embody community and his teachings continue to inspire how I move through the world to this day.

Thank you to Dr. Astrid Brousselle for being an encouraging and thoughtful supervisor throughout this project and Dr. Jill Chouinard for being an inquisitive second reader. I also want to thank the MACD 2018 cohort for being an all-around thoughtful and compassionate group of people. I look forward to when and wherever our paths cross again, despite all of us being peppered across the country.

Many heartfelt thanks to all of my CSS colleagues for supporting me throughout this program. You are an inspiring, dedicated, talented and loving group of people that I am privileged to work alongside. A special thanks to Sharon Yeo for overseeing and supporting this project from start to finish.

Lastly, I want to thank all of my family and friends who have supported me for the past three years of full-time work and study. Most of all, thank you to my partner Caleb for being the person beside me every single day over the past three years encouraging me forward.

Executive Summary

Introduction

Background

This research project completed an evaluation capacity assessment with the Immigration and Settlement Service (ISS) of Catholic Social Services (CSS), a newcomer-serving social service nonprofit based in Edmonton, Alberta. The goal of the project was to understand the organization's evaluation capacity, including facilitators and barriers to evaluation capacity, so that informed recommendations could be made to build the evaluation capacity of the organization. This research is situated within the environment of public accountability and governing for results that shapes nonprofit-government relations in Canada, with the central focus of the project being on evaluation as a mechanism of both accountability and organizational learning.

Objectives

The purpose of the study is to understand ISS' current evaluation capacity by identifying barriers and facilitators to building evaluation capacity, and to identify tangible actions the organization can take to build internal evaluation capacity.

The primary research question is:

- What is Immigration and Settlement Services' current evaluation capacity, and what are the barriers and facilitators to build evaluation capacity?

The secondary research questions include:

- What are managers, team leaders and staff's experience with evaluation?
- How can ISS develop a culture of evaluation?

Literature Review

A literature review was completed on the topics of nonprofit-government relations, accountability, evaluation and performance measurement, evaluation capacity building and cultures of evaluation, and evaluation practices within the nonprofit sector. A conceptual framework was created to visualize the key findings from the literature review and guide how the data was analyzed and integrated with the existing body of literature.

Methodology and Methods

Methodology

The project is based on a mixed methods case study design (Creswell and Plano Clark, 2017). According to Yin (2009), a case study is a methodological approach to research in which the researcher investigates a specific "phenomenon in depth within its real-life context" (p. 18). The case study approach is to better understand how evaluation as a phenomenon, including data collection practices, staff experience and perception, and evaluation uses, contribute to a culture of evaluation in the organization. The case study uses a mixed method design in which quantitative data is gathered first, followed by a qualitative method that builds on the analysis and findings of the quantitative data. This design is described by Creswell and Plano Clark (2017) as a sequential explanatory design (p. 77).

Methods

A survey was distributed to all staff of ISS to gather quantitative data, which was followed by small group discussions with staff to present the survey findings and further discuss the evaluation practices of the agency and explore opportunities to build evaluation capacity as a group. The survey used was an adaptation of the Evaluation Capacity Assessment Instrument, which is a validated and empirical evaluation assessment tool specifically for nonprofit organizations (Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry and Balcazar, 2013). The tool measures Individual, Organizational and Evaluation Capacity Outcomes as factors that contribute to evaluation capacity. Three small group discussions were then held with a purposeful sample of staff at the Manager, Team Leader and Coordinator level to build on and contextualize the results of the survey.

The survey had 41 participants from Immigration and Settlement Service, out of 115 full-time and part-time regular staff, which is a response rate of 36%. Three small group discussions were held and attended remotely by 13 staff.

Analysis

The data from the survey was analyzed by calculating the mean scores for each statement within the Individual, Organizational and Evaluation Capacity Outcomes factors, including the creation of bar charts that demonstrate the lowest and highest scoring statements within each factor.

The small group discussions were recorded, transcribed and then examined through multiple close readings. Key quotes from the transcriptions were arranged in categories that addressed the research questions.

The findings from both the survey and discussions were then compiled in a table to integrate the findings from both methods. Lastly, the major findings from the previous analysis activities were then arranged visually into the conceptual framework in an effort to bridge the findings into recommendations to build evaluation capacity.

Key Findings

1. **Staff are aware of the benefits** of evaluation and **are motivated to learn** more about the topic.
2. **Staff have varied competence for evaluation**, including many who have attended outcome measurement training through the organization, a funder or post-secondary institutions. There is a strong foundation of evaluation skills and practice at ISS, particularly in data collection methods, although staff skills could be updated, as the recency of evaluation specific training ranged from current to more than a decade ago.
3. The organization has some resources in place for evaluation, such as the provision of interpretation to facilitate client participation in evaluation, but other resources are

noted as significant **barriers to evaluation capacity,** including **lack of time,** designated **evaluation staff** or department, and evaluation leadership and **guidance from a major funder.**

4. The organization predominantly uses and communicates **evaluation as a tool of accountability**, which limits its uses for other key organization functions, such as organizational learning, staff engagement and social innovation. The organization would benefit from balancing the intended uses of evaluation to maximize its utility across varied stakeholders, including frontline staff.
5. The organization would benefit from undertaking evaluation capacity building interventions to promote a culture of evaluation, specifically to **engage all staff in the evaluation process, communicate evaluation findings and results in a timely manner, and implement varied approaches to evaluation**. Additional approaches include collaborative and culturally responsive approaches, particularly in light of the challenges associated with the population ISS serves, such as differences of language, culture, and power dynamics.

RECOMMENDATIONS

1. *Invest in evaluation* through an internal evaluator or staff who can ensure the continuity of evaluation practices within the service and spearhead evaluation capacity building opportunities with staff.
2. *Balance the intended uses of evaluation* between accountability and other key organizational functions of equal value, including organizational learning, staff engagement, capacity building, and social innovation.
3. *Explore collaborative and culturally responsive approaches* to evaluation as a means to build evaluation capacity, engage staff, increase evaluation utility, and ensure evaluation is responsive to the cultural context of clients.

Table of Contents

Acknowledgements 2

Executive Summary 3

Table of Contents 6

List of Tables 8

List of Figures 8

1.0 Introduction 9

1.1 Background 9

1.2 Project Client 10

1.3 Defining the Problem 12

1.4 Project Objectives and Research Questions 13

1.5 Positionality of Principal Researcher 13

1.6 Organization of Report 14

2.0 Literature Review 15

2.1 Government-Nonprofit Relationship 15

2.2 Accountability 16

2.3 Evaluation and Performance Measurement 17

2.4 Evaluation Capacity and Culture of Evaluation 20

2.5 Evaluation Practices and Capacity of Nonprofits 23

2.7 Gaps in the Literature 25

2.8 Conceptual Framework 25

3.0 Methodology and Methods 26

3.1 Methodology 26

3.2 Methods 27

3.3 Data Analysis 32

3.4 Ethical approval 33

3.5 Research credibility and generalizability of results 33

4.0 Findings 36

4.1 Introduction 36

4.2 Evaluation Capacity Assessment Instrument 36

4.3 Small Group Discussions 42

4.4 Integration of the Methods 49

5.0 Discussion and Analysis 53

5.1 Evaluation Capacity at ISS 53

5.1.1 Individual Factors 54
5.1.2 Organizational Factors 55
5.1.3 Evaluation Capacity Outcomes 57
5.2 Culture of Evaluation 58
5.3 Additional Implications 58
6.0 Recommendations 60
6.1 Introduction 60
6.2 Recommendations 60
6.2.1 Invest in Evaluation 60
6.2.2 Balance Intended Uses of Evaluation 60
6.2.3 Explore Collaborative and Culturally Responsive Approaches 60
7.0 Conclusion 62
References 63
Appendices 72

List of Tables

Table 1. Mixed Method Case Study 27
Table 2. Position of Respondents and Actual Staff Count 29
Table 3. Tenure with Agency of Respondents 30
Table 4. Position of Small Group Discussion Participants 31
Table 5. Inventory of Current Evaluation Practices 43
Table 6. Joint Display of ECAI and Small Group Discussion Findings 50

List of Figures

Figure 1. Conceptual Framework of Evaluation Capacity 25
Figure 2. Bar Chart of Awareness Responses 37
Figure 3. Bar Chart of Motivation Responses 38
Figure 4. Bar Chart of Competency Responses 38
Figure 5. Bar Chart of Leadership Responses 39
Figure 6. Bar Chart of Learning Climate Responses 40
Figure 7. Bar Chart of Resources 41
Figure 8. Bar Chart of Mainstreaming 42
Figure 9. Bar Chart of Use of Findings 42
Figure 10. Evaluation Capacity at ISS within Conceptual Framework 53

1.0 Introduction

This project will complete an evaluation capacity assessment of the Immigration and Settlement Service of Catholic Social Services. The goal of this project is to explore the evaluation capacity of the organization through a mixed methods case study and provide the organization with recommendations to build evaluation capacity. This section will provide a background to the project, describe the project client and their work, define the problem, describe the project objectives and research questions, and establish the positionality of the researcher.

1.1 Background

Governments in Canada and around the world are increasingly contracting nonprofits to deliver public services to reduce the cost of services in times of fiscal restraint and to increase efficiency and effectiveness, leading to both privatization of service delivery and increased demands for accountability (Brinkerhoff, 2002; Despard, 2016; Smith, 2016). As a result, nonprofit organizations have become a common provider of public services, such as health care and social services (Bach-Mortensen and Montgomery, 2018). At the same time, public institutions continue to adopt service delivery strategies and approaches that are characterized by accountability and governing for results, such as New Public Management (NPM) approaches (Chouinard, 2013). Poister et al. (2014) refer to the rise of accountability as the "government rationality movement," which aims to make government administration more objective with evidence-based strategies (p. 4). Accountability serves a function of control; public funders and subsequently taxpayers are confident that their financial resources are being used responsibly through accountability practices (Gill, 2010). Performance measurement and evaluation are instruments of accountability and evidence-based practice that have been enshrined in Canadian federal institutions through the Treasury Board Secretariat's (TBS) Policy on Results (TBS, 2016a) and subsequently the Directive on Results (TBS, 2016b). Governments at all levels require performance measures and evaluation within their accountability frameworks and the organizations they fund.

Within this context, nonprofits have struggled to meet the demands for accountability placed on them (Bach-Mortensen and Mongomery, 2018; Carman, 2007; Carman and Fredericks, 2010; Despard, 2016; Stoecker, 2007). In North America, nonprofits that receive government funding are often required to report on performance measures and evaluate their services to demonstrate that they have met service targets outlined in contractual agreements and that services have delivered their intended outcomes (Carman and Fredericks, 2010). Performance measurement and program evaluation are means used by governments and funded agencies to demonstrate public accountability, along with other management practices including accreditation, financial audits, performance measurement of individual personnel, and other forms of reporting (Carman, 2007). Nonprofits face distinct challenges to evaluate program outcomes and measure their performance, which include lack of staff, training, access to sophisticated software, time and financial resources (Despard, 2016). As such, nonprofits can benefit from empirical assessments of their evaluation capacity (Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry and Balcazar, 2013) and applying tailored interventions to build evaluation capacity (Preskill and Boyle, 2008).

This project will complete an evaluation capacity assessment with the Immigration and Settlement Service of Catholic Social Services, a newcomer-serving social service nonprofit based in Edmonton, Alberta. The goal of the project is to understand the organization's evaluation capacity, including facilitators and barriers to evaluation capacity, so that informed recommendations can be made to the build evaluation capacity of the organization. This research is situated within the aforementioned environment of public accountability and governing for results that shapes nonprofit-government relations in Canada, with the central focus of the project being on evaluation as a mechanism of both accountability and organizational learning.

1.2 Project Client

Catholic Social Services (CSS) was founded in 1961, has grown to employ over 1,800 staff and 700 volunteers annually, and operates in 12 municipalities in Northern Alberta (CSS, 2019). The agency provides social support to vulnerable Albertans in three distinct service areas: Community, Family and Child Service, Community Outreach and Disability Service, and Immigration and Settlement Service (ISS). The agency's mission is to "care for and bring hope to those in need with humility, compassion and respect" (CSS, 2020). The annual operating budget of the Agency in the 2019 - 2020 fiscal year totaled $102 million CAD (CSS, 2020).

The ISS of CSS has been serving newcomers in Alberta for over 58 years and employs 115 full-time and part-time staff, in addition to fluctuating relief staff, in Edmonton and Red Deer. The staff employed by ISS often reflect the clients they serve, which is evidenced by staff in Edmonton and Red Deer speaking 70 languages. Settlement services are a sub-sector of social services that assist newcomers with all aspects of daily life that allow them to settle in their new country (Praznik and Shields, 2018). The primary services delivered are information and orientation on the topics of government documentation, educational institutions and systems, healthcare, community connections, rights and responsibilities, housing, and language testing and training. ISS serves approximately 14,000 newcomer clients per year in Edmonton and Red Deer. ISS' two primary funders are Immigration, Refugees and Citizenship Canada and the Government of Alberta. ISS also receives funding from Family, Community and Child Services, which is operated by the City of Edmonton, and other special grants. Lastly, CSS operates the Sign of Hope charity subsidiary of the agency, which provides funding to many emergent programs that have not secured stable public funding.

ISS offers a suite of federally funded programs to meet the needs of different classes of permanent residents. First, ISS operates the Language Assessment, Referral and Counselling Centre (LARCC), which administers the Canadian Language Benchmark exam. The Canadian Language Benchmark exam score is a requirement to be placed in the federally funded Language Instruction for Newcomers to Canada (LINC) language program, which permanent residents can attend at no cost. Second, ISS delivers the Resettlement Assistance Program, which provides settlement services to Government Assisted Refugees (GARs) for their first six weeks in Canada, including arrival at the airport and fourteen days of temporary housing. After six weeks, GARs are then transferred internally to the Complex Case Management Program, which provides case management services for the remainder of GAR clients' first eighteen months in Canada. The language assessment program and RAP are cornerstone programs of ISS because the federal government only funds one language assessment centre and one RAP contribution agreement per urban center. Therefore, CSS is the only organization in Edmonton that operates these federal programs.

ISS also delivers the Newcomer Settlement Program, Enhanced Case Management Program, Orientation, Volunteer Programs, Cross-Cultural Counselling, and Community Support Worker Program, all of which are federally funded. Newcomer Settlement Program is a catchall settlement program for all classes of permanent residents, including economic immigrants, family class immigrants, Privately Sponsored Refugees, and GARs after they have been in Canada for more than one year. Orientation provides a five-day course, delivered once a week for five weeks, to recently arrived GARs on the topics of rights and responsibilities, child and adult education, healthcare, child welfare, weather and driving. The course is internally referred to as "a crash course on life in Canada" for GAR clients. Volunteer Programs offer community connection opportunities for established Canadians to support newcomer settlement, through English language exchanges and formal friendship matching, and opportunities for newcomers to gain Canadian work experience through volunteer placements. Cross-Cultural Counselling provides mental health support to newcomers through individual and family cross-cultural counselling. Community Support Workers provide direct support services for newcomers that facilitate their access to settlement and mainstream services, most often through interpretation.

The Government of Alberta funds two ISS programs, including Language and Vocational Assessment and the Community Support Services Program. Language and Vocational Assessment, which is also part of the LARCC, provides newcomers with individual support and counselling related to language learning options, career pathways, and Canadian education institutions. The Community Support Services Program provides settlement services to foreign nationals who are not permanent residents, including Temporary Foreign Workers and International Students, as they are not eligible for federally funded settlement programs. The Government of Alberta also recently funded the Family Wellness and Community Enhancement Program until September 1, 2020, which worked with newcomer families to address issues of family discord as a result of reunification.

ISS receives funding from the City of Edmonton's Family and Community Support Services for the Cross-Cultural Counselling, Youth programs for GARs, and the Language and Community Enrichment (LACE) program, which delivers English conversation circles at 18 Edmonton Public Library locations. Sign of Hope provides small amounts of supplementary funding to the majority of ISS' programs, most notably the following programs: Intercultural Education, Parenting in Two Cultures and Outdoor Wellness.

ISS in its entirety will be the organizational focus of this research project. The objective is to understand their collective evaluation capacity across their varied programs, rather than evaluate a single program. The underlying goal of the project is to support the ISS to strengthen evaluation efforts and take steps towards developing a culture of evaluation. Ultimately, strengthening evaluation practices and developing a culture of evaluation would support ISS to continue to be competitive in a challenging funding climate, improve current programs and continue to fill service gaps for clients, and engage staff.

1.3 Defining the Problem

ISS needs to pursue numerous evaluation activities, as public accountability activities, given that it receives the majority of its funding from various levels of government. Despite its myriad of publicly funded programs, ISS does not have an evaluation department, committee or individual staff member specifically designated and trained to evaluate its programs. Individual programs all complete evaluative activities, such as surveys and client focus groups for accountability purposes, but there is little capacity to compile, analyze or use the data gathered in a systematic and meaningful way, or communicate findings to staff. Furthermore, both the federal and provincial governments are operating from perspectives of fiscal restraint, which creates increased competition among agencies to deliver cost-per-client services. Evaluation capacity building would support ISS to determine the effectiveness of their programs, support ongoing improvement and innovation to service delivery and develop a culture of learning. A culture of learning would increase their overall adaptability in the current environment of political, social and fiscal uncertainty. Evaluation would also support ISS to be more responsive to the changing needs of their clients, by listening to, documenting and strategically incorporating client voices into programs and service delivery strategies, ultimately influencing positive client outcomes.

All of ISS' major funders, including IRCC, the Government of Alberta and Family and Community Support Services, announced funding decreases in the April 1, 2020 fiscal. Immigration, Refugees and Citizenship Canada funds provinces and regions based on permanent resident arrivals to the designated region. Historically, Alberta's economy experienced decades of sustained economic growth resulting in a rich pool of employment opportunities, which subsequently attracted immigrants to settle in the province. However, Alberta's economy has been in decline since 2016, recently compounded by the COVID-19 pandemic, which has resulted in fewer immigrant arrivals in the province. As a result, settlement agencies operating in Alberta faced a decrease in federal funding in the spring of 2020 for the first time in many decades. In addition, the current provincial government has also announced major cuts to public services, most notably to education, social services and health care. To date, the Province of Alberta has not announced cuts to any of the programs it funds at ISS, but it is a possibility during the next three years of the current provincial government. The province has also decreased its funding to municipalities, which impacts funding to Family and Community Support Services through the City of Edmonton. As a result, agencies in Alberta, including ISS, are dealing with a smaller pool of funding for their programs from all levels of government, and therefore need to prove the value and effectiveness of the services they offer. As noted previously, evaluation can be a way to judge the merit of services offered and increase organizational learning, adaptability and innovation, which is increasingly important in the current fiscal environment.

1.4 Project Objectives and Research Questions

The purpose of the study is to understand ISS' current evaluation capacity by identifying barriers and facilitators to building evaluation capacity, and to identify tangible actions the organization can take to build internal evaluation capacity.

The primary research question is:

- What is Immigration and Settlement Services' current evaluation capacity, and what are the challenges and opportunities to build evaluation capacity?

The secondary research questions include:

- What are managers, team leaders and staff's experience with evaluation?
- How can ISS develop a culture of evaluation?

An assumption embedded within this project is that improved evaluation capacity will support ISS to develop a culture of learning, improve evaluation practices to enable them to evaluate the effectiveness of their programs, and identify opportunities for service delivery improvement and innovation.

1.5 Positionality of Principal Researcher

It is worth noting that I have a vested interest in the results of this research, given my employment within the organizational focus of this project. Furthermore, prior to undertaking this research, I was already embedded in the organization and had my own experiences and observations related to the evaluation practices of the organization that have shaped how I have approached and understood the project. Although participatory action research is not the methodological approach of the project, it is important to note my positionality and personal relationship to the research in an effort to name biases rather than ignore or omit them (Guishard, 2008; Muhammad, Wallerstein, Sussman, Avila, Belone and Duran, 2015; Patton, 1999).

I am a settler on Treaty 6 Territory, the traditional territory of the Cree, Saulteaux, Nakota, and Dene peoples. I have lived and worked in Edmonton, Alberta for the past five years in the newcomer settlement sector. I began the Master of Arts in Community Development (MACD) at the University of Victoria in May 2018 and began a new role at ISS shortly afterwards in July 2018. My new role as Special Projects Coordinator required me to complete focus groups with clients and support the development, distribution and analysis of program surveys across ISS. I also frequently draft funding proposals for various programs for all levels of government, all of which have sections related to evaluation and outcome measurement. Combined, these duties contribute to a significant portion of my workload. I had very little knowledge of evaluation when I started in this new role, apart from undergraduate courses in social science research methods. This led me to begin reading literature on evaluation to strengthen my ability to fulfill the evaluation aspects of my job.

At the same time, I completed courses in the MACD program, including community-based research and ethics, leadership, and project design, implementation and evaluation. As a result of these courses and through informal learning, I began to see room for growth in the Agency's evaluation practices. Dr. Astrid Brousselle recommended that I take the course *ADMN 537: Program Evaluation and Performance Measuremen*t at the School of Public Administration to increase my knowledge of program evaluation and concretize my research project after she had approved my research proposal and agreed to be my supervisor. I joined an Evaluation Communities of Practice with staff of other social service providers in Edmonton, which I have attended once a month since September 2019.

Taken together, this project has been a learning process for me as I further increase my knowledge of the field of evaluation, complete my project and fulfill my current role at ISS. My goal is to collectively increase the evaluation knowledge and competency of my workplace, which will lead to improved organizational learning and support program development that will positively impact newcomers settling in Edmonton. At the same time, I am now critical of the traditional evaluation methods used by many evaluators in the social service sector, particularly the settlement sector. I often question the validity of evaluation methods used with clients who are unfamiliar with surveys, focus groups, and other methods, and who either may not speak an official language or read or write in their first language. It is my hope that the project can be a catalyst for further conversations about how we can explore different methods that are more culturally appropriate in the context of our work and redistribute power to clients and frontline staff in the evaluation process.

1.6 Organization of Report

This report will be organized into the following sections: a review of the relevant literature and conceptual framework for the research; methodology and methods undertaken; research findings; discussion and analysis of research findings, and lastly, recommendations and conclusion.

2.0 Literature Review

This literature review covers the following topics: government-nonprofit relationships, accountability, evaluation and performance measurement, evaluation capacity, nonprofit evaluation practices and capacity, evaluation capacity building, and culture of evaluation.

2.1 Government-Nonprofit Relationship

According to Anheier (2005), the need for social services will always outweigh the ability of nonprofits to secure private donations, resulting in nonprofits seeking reliable funding streams, most often through government sources (p. 282). Governments fund public sector nonprofits to provide essential public services, such as health care and social services (Quarter and Mook, 2009). An underlying assumption in their relationship is that governments are superior administrators while nonprofits are better at service delivery due to their connection and integration in local communities (Brinkerhoff, 2002; Quarter and Mook, 2009; Salamon and Toepler, 2015). Furthermore, governments are able to reduce costs of service provision by contracting nonprofits, who often operate with supplementary contributions from private donations and volunteers (Knutsen, 2017; Smith, 2016).

Governments and nonprofits work in partnerships that present benefits and challenges for both parties (Brinkerhoff, 2002; Knutsen, 2017; Salamon and Toepler, 2015). The benefits identified in the literature include increased pluralism and civic activism, reduced administrative burden, increased efficiency through competition between agencies, collaborative potential around emergent issues, and more tailored services as a result of connections to local communities that nonprofits leverage (Brinkerhoff, 2002; Knutsen, 2017; Salamon and Toepler, 2015). In addition to receiving direct payments for fee-for-service contracts, nonprofits experience the benefits of tax exemption, tax deductions for donators and preferential regulatory treatment (Anheier, 2005, p. 283). The negative consequences of government-nonprofit relationships primarily relate to the imbalance of power between governments and nonprofits that result in nonprofits often changing activities and strategies to maximize government contracting at the expense of their social mission (Ebrahim, 2005; Knutsen, 2017). Smith and Lipsky (1993, as cited in Anheier, 2005, p. 287) argue that governments and nonprofits participate in "mutually dependent" but unequal relationships in which governments act as a hegemonic leader that subordinates nonprofits.

Anheier (2005) presents government and nonprofit relationships as: (1) substitute and supplementary, (2) complementary, or (3) adversary. Nonprofits may supplement or substitute services when there is an identified gap in public services being provided for vulnerable people. Governments and nonprofits demonstrate a complementary relationship when governments fund nonprofits through stable funding, namely through tax income, and in turn nonprofits provide effective and responsive services in local communities. Lastly, nonprofits can at times side with minority perspectives on issues, such as those represented in social movements or protests, creating a political conflict that results in an adversarial relationship. According to Young (2000, as cited in Anheier, 2005, p. 284), all three classifications of government-nonprofit relationships may be present and at play at a given time.

Also significant, Najam (2000) identifies four C's model of nongovernmental organization (NGO) and government relations, which include cooperation, co-adaptation, complementarity and confrontation. Najam notes the possibility of a relationship of non-engagement, where select NGOs may choose not to engage with government either entirely or on specific topics (p. 384). Their model focuses on the topic of the preferred ends or goals, as well as the preferred means or strategies of each party, all of which can shape the relationship between government and nonprofits.

Governments and nonprofits are logically coupled due to demands on governments to meet public needs and the ability of the nonprofit sector to respond to emergent societal needs with urgency and more efficiency than governments. There are benefits to nonprofits delivering public services in place of direct government services, such as healthcare and social services, including the use of voluntary labour, tax benefits and exemption, and connection and responsiveness to local communities. However, nonprofits have a tendency to alter their activities to maximize public contracting, which can compromise their social missions.

2.2 Accountability

Since the 1960s, governments and publicly funded organizations have been impacted by the rise of New Public Management (NPM), an approach to governance that strives for increased efficiency, accountability and managing for results (Chouinard, 2013). Given the nonprofit sector's close relationship with governments, they have also been impacted by the growing emphasis on accountability, results and efficiency (Whitaker, Altman-Sauer and Henderson, 2004). More recently, the approach has shifted from NPM to other approaches in the field of public administration, including New Political Governance, but governments have retained the focus on accountability, efficiency and evidence-based decision making brought about by NPM (Chouinard and Milley, 2013; McDavid, Huse and Hawthorn, 2017). 'Accountability' is seen as an indisputable good that needs to be pursued and continually improved as it provides assurance that taxpayer dollars are being spent wisely and that public interest and priorities are being respected (McDavid, Huse and Hawthorn, 2017; Whitaker, Altman-Sauer and Henderson, 2004). McDavid, Huse and Hawthorn (2017) present the idea that there is a common assumption that public tax dollars will be wasted and the politicians and public servants will act in self-interested ways if strong accountability systems are not in place (p. 429).

Chouinard (2013) questions the prominence of accountability rhetoric within public management discourse, particularly the argument that more accountability is always desired. She argues that the term accountability has shifted from meaning the presence of shared responsibility to encompass "control, regulation, and compliance" (p. 239). Furthermore, she questions how much accountability is considered enough and whose interests does ever-increasing accountability serve. Ebrahim (2005) echoes this sentiment by stating that exhaustive accountability requirements for nonprofits actually detract from their social mission and negatively impact their ability to carry out their work, particularly where instances of accountability and resource dependence are coupled. They further state that the accountability requirements placed on nonprofits actively inhibit organizational learning and create cultures of compliance and risk aversion, which limits the ability of nonprofit organizations to experiment with programming, potentially preventing them from discovering social innovations to complex societal problems. These arguments are not meant to imply that organizations should not be accountable for public funds, but rather that the continually growing administrative burdens placed on nonprofits in the name of 'accountability' negatively impact their capacity to actually fulfill their missions and find solutions to the social issues they are meant to address.

2.3 Evaluation and Performance Measurement

2.3.1 Evaluation

Performance measurement and evaluation have become central to systems of accountability. There are varied definitions and approaches to evaluation. The Canadian Evaluation Society (2015) defines evaluation as the "systematic assessment of the design, implementation or the results of an initiative for the purposes of learning or decision-making." According to McDavid, Huse and Hawthorn (2017), evaluation is an organizational activity that gathers information to test the effectiveness of policies and programs and areas of improvement through hypotheses and research designs (p. 3). Wholey, Hatry and Newcomer (2010) offer a comparable definition by stating that program evaluation is the systematic analysis of program results and whether the program of focus caused the results. Chouinard and Cram (2020) further add that evaluation, in addition to addressing questions of program improvement and effectiveness, serves the purpose of strategic decision-making (p. 1). Evaluation is situated within the literature as both an organizational practice and cultural orientation towards evidence-based decision-making, strategic planning and social innovation (Chelimsky, 1997).

McDavid, Huse and Hawthorn (2017) describe the steps to conduct and report an evaluation, which include: understanding the context and needs of the evaluation, such as determining the research questions; developing an evaluation plan and the data collection instruments (methods); collecting lines of evidence through pre-determined methods; analyzing data as it pertains to the research questions; writing and reviewing an evaluation report; and, sharing the report and findings with applicable stakeholders (p. 29). These steps vary depending on the approach to evaluation, as collaborative and culturally responsive approaches heavily emphasize stakeholder and participant consultation throughout the evaluative process (Cousins, 2019; Chouinard, 2020). Regardless of the difference in approaches, evaluation is a process that includes the use of methods to collect lines of evidence that answer fundamental program, policy or intervention questions, including impact on participant lives and communities, implementation strategy and improvement, service gaps and solutions, among others.

Evaluation can be distinguished from other accountability fields, such as accreditation and accounting, because of its emphasis on testing causality between policies and programs and observed outcomes (McDavid, Huse and Hawthorn, 2017, p. 12) and its ability to support organizational learning and development. Evaluation is most often used to determine the connection between social interventions, such as policies and programs, and observed outcomes in populations and for organizational improvement and growth. Furthermore, proponents of collaborative approaches to evaluation, discussed later in this section, position evaluation as a tool to further social justice because of its ability to measure and improve the social impact of interventions and empower, listen to, and amplify the voices of typically marginalized populations who may be participants in programs and communities being evaluated (Chouinard, 2020; Cousins, 2019).

Chelimsky (1997) provides the argument that evaluation falls under three primary categories and corresponding purposes: improvement, accountability and knowledge. Gill (2010) adds that evaluation is for results, learning and change making. Both Chelimksy and Gill's categories reflect that evaluation is for accountability and demonstrating results, as well as for organizational improvement, growth and innovation. Therefore, evaluation is a multifaceted practice that organizations can use for multiple purposes, including for accountability and a tool for organizational learning and development, which cannot be attributed to other fields of organizational accountability, such as accounting.

Evaluation tends to be split into three key areas: summative evaluation, which determines if a program or policy has been effective in achieving its intended outcomes, and formative, which aims to understand how the program was delivered and how it could be improved (McDavid, Huse and Hawthorn, 2017). Patton (2011) adds developmental evaluation to the literature, which is an approach to evaluation that incorporates systems thinking and complexity to respond to the ongoing developments within systems and programs by gathering real-time data and making informed decisions in a dynamic environment (p. 1). Program evaluations can combine aspects of summative, formative and developmental evaluation by answering questions related to program outcomes and more technical aspects of how programs were delivered and how delivery can be improved.

The field of evaluation also has a large body of approaches, including but not limited to: utilization focused evaluation (Patton 2012), and collaborative approaches (Cousins, 2019), which includes empowerment evaluation (Fetterman, 2002), culturally-responsive approaches (Chouinard and Cram, 2020), among others. The differences in approaches can be marked by the epistemological assumptions related to the evaluator, program participants, and wider stakeholders, preferred methods, rationale, context, what information is considered valid and defensible, among others. According to McDavid, Huse and Hawthorn (2017), different approaches to evaluation can and should be adopted based on the context of the evaluation, and as such evaluators need to be well versed in multiple approaches and data collection methods. They state that evaluations being done in a high-stakes and accountability focused context, such as a high profile government program being at risk of discontinuation, requires a more traditional, highly rigorous, and scientific approach to evaluation. However, if the primary objective of the evaluation is organizational learning, then a collaborative approach may be more applicable (Cousins, 2019; McDavid, Huse and Hawthorn, 2017).

2.3.2 Performance Measurement

Performance measurement is the systematic collection of quantitative data related to policy or program activities that are used to inform and advance organizational goals (Poister et al., 2014, p. 7). According to Poister (2010), performance measures include outcomes, cost-effectiveness, outputs, quality, and customer satisfaction (p. 101). Outcomes are considered the most important measure of performance monitoring systems, followed closely by outputs

(Poister, 2010). Outputs and outcomes as standalone features of a performance monitoring framework, while useful, cannot draw causal conclusions about the effectiveness or impact of a program, which is why evaluation designs are necessary for more in depth analysis of a program (Poister, 2010, p. 104). Performance measures tend to focus on the quantitative and normative realities of programs, such as how many individuals were served or how many individuals demonstrated a desired program outcome, whereas evaluation studies whether the program activities were effective, its importance and future relevance, whether it is well designed and well implemented, and participants' contextual experiences in the program. Taken together, program evaluation provides a more nuanced understanding of a program or policy.

Poister et al (2014) situate performance measurement and program evaluation as organizational reform cousins, both of which provide leaders with insight into the accomplishments of the organization and inform future strategic planning (p. 27). Program evaluation is more resource intensive to conduct than gathering performance measures, therefore in times of fiscal restraint governments often lean more heavily on performance measures to make strategic decisions, despite the fact that program evaluation provides deeper understanding of the social outcomes of policies and programs (McDavid, Huse and Hawthorn, 2017). Furthermore, evaluation provides opportunities to listen to and understand the social context of program participants and communities, which is not possible with performance measures alone. McDavid, Huse and Hawthorn (2017) argue that heavily relying on performance measures at the expense of program evaluation has negative impacts on organizational learning, overall effectiveness and staff morale (p. 5).

2.3.3 Collaborative Approaches to Evaluation

Cousins (2019) loosely defines collaborative approaches to evaluation as any evaluation in which trained evaluators work alongside non-evaluators to "coproduce evaluative knowledge" (p. 5) in the "planning, implementing, and disseminating" of evaluative knowledge (p. 7). The literature notes multiple collaborative approaches to evaluation that support organizations and communities to evaluate their programs and build evaluation capacity through collective involvement in the evaluation process (Fetterman, 2001; Chouinard, 2013; Acree and Chouinard, 2019; Chouinard and Cram, 2020; Janzen, Ochocka, Turner, Cook, Franklin and Deichert, 2017; Suarez-Balcazar and Harper, 2004). Collaborative approaches to evaluation have pragmatic, political and philosophical justifications related to the practical ability of community members to offer more in-depth understanding on a program context, sharing of power and displacing traditional epistemological practices with academic as expert, and to continue exploring the complexities in which programs operate (Cousins, 2019, p. 11 - 12).

Participatory evaluation (Suarez-Balcazar and Harper, 2004) is rooted in the tradition of participatory action research and centralizes the concept of active and ongoing involvement of stakeholders in the evaluation process and shared control over critical decisions related to the evaluation and its implementation (p. 2). They argue that participatory approaches to evaluation are both useful and ethical when evaluating programs in communities that have historically been marginalized and oppressed. Participatory approaches have been described by

Cousins and Whitmore (1998) as “a three-pronged” activity which includes investigation, education and action (p. 8).

Proponents of culturally responsive approaches to evaluation posit that culture is fundamental to the context of every evaluation, not only in terms of program participants and communities but the ways in which the evaluator worldview impacts how information is collected, understood, and shared (Chouinard and Cram, 2020). Chouinard and Cram have identified nine dimensions of culture that impact evaluation, including: epistemological, ecological, methodological, political, personal, relational, institutional, axiological and ontological dimensions. Fundamentally, culturally responsive approaches challenge evaluators to consider how their worldview and cultural biases impact evaluation and argue that the equitable inclusion of community members in evaluation not only enriches evaluation findings and use but is ethically imperative.

It is important to note that collaborative and participatory approaches to research and evaluation are not without criticism (Cornwall and Jewkes, 1995; Cornwall, 2008). Power differentials between researchers and community members are enduring, despite the best efforts of researchers to work from an anti-oppressive lens (Cornwall, 2008). As well, community developers and researchers often lump marginalized communities into monolithic groups, and erase the differences between individual members. Cornwall and Jewkes (1995) further problematize participatory evaluation stating it could have negative impacts on program participants, communities may be skeptical or resistant to research or evaluation projects, and that the equitable participation of community members across race, gender, financial standing, among other factors is unlikely (Cornwall and Jewkes, 1995).

2.4 Evaluation Capacity and Culture of Evaluation

For Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry and Balcazar (2013), evaluation capacity is the collection of organizational and individual factors present in an organization or system that facilitates evaluation and its use (p. 192). In their synthesized model for nonprofits, the Evaluation Capacity Assessment Instrument (ECAI), evaluation capacity can be measured in the following areas:

- Individual factors, including awareness of evaluation and its benefits, motivation to pursue evaluation, and formal competence and skill to conduct evaluations;
- Organizational factors, including leadership’s support of evaluation, evaluative thinking and learning climate, and the resources to conduct evaluation, such as paid evaluation staff and technology; and,
- Evaluation outcomes, which include how evaluation has been mainstreamed into the day-to-day practices of the organization, including in formal policy, and the use of evaluation findings (p. 192).

Similarly, Cousins, Goh, Elliot and Bourgeois (2014) group evaluation capacity into the formal capacity *to do* evaluation and the capacity *to use* evaluation. The capacity to do evaluation includes the professional competencies the individual or organization has to effectively evaluate their programs, including the formal education and training, data collection tools and methods, data analysis and interpretation, and written reports, and the informal skills that facilitate evaluation, such as relationship building and conflict resolution (p. 16). The capacity to use evaluation includes the ability of the organization to integrate evaluation knowledge into the day-to-day operations and strategies of the organization, which includes the organization's integration of evaluation findings in program planning and improvement, collective learning, discovery and reflection, and symbolic uses, such as affirmation of program value and staff effort (p. 17).

According to Gagnon, Aubry, Cousins, Goh and Elliot (2018), evaluation capacity is the intersection of an organization’s ability to do evaluation, to use evaluation and to pursue organizational learning. In a recent empirical study, they measured the validity of constructs often attached to evaluation capacity, including: organizational learning, support structures, capacity to do evaluation, stakeholder participation, capacity to use evaluation, and mediating factors. They found that these constructs are valid measures to determine the evaluation capacity of an organization.

Bourgeois and Cousins (2008) provide a framework for organizational evaluation capacity where organizations have low capacity, developing capacity, intermediate capacity or exemplary capacity. They argue that it is necessary to identify the evaluation capacity of an organization before pursuing evaluation capacity building approaches. According to their framework, organizations exhibit low capacity when they typically contract evaluators for evaluations of individual programs for funder requirements, have little technical skills to conduct evaluation, evaluations conducted lack rigor or defensibility, and there is little or no evidence of evaluations being used to inform program strategy or decision-making. Organizations exhibit exemplary evaluation capacity when they consider evaluative enquiry as a core management function, it is fully integrated into decision-making, and all staff demonstrate results-management and evidence-informed practice.

Volkov and King (2007) have developed an evaluation capacity building checklist for organizations to inventory their internal evaluation capacity and key areas of development. The checklist includes the following broad themes:

- Organizational context, including positive evaluative culture internally and external environment, such as funder accountability directives;
- Evaluation capacity building structures, such as inclusion of evaluation in strategic plans, the infrastructure to support evaluation activities and communication, integration of evaluation in day-to-day operations, and evaluation-related learning opportunities, such as peer support; and,
- Evaluation resources, including the provision and expansion of resources and systems that provide support to evaluation, such as long-term financial support.

Preskill and Boyle (2008a) define evaluation capacity building as the intentional design and implementation of teaching and learning strategies about evaluation practice, with the ultimate goal being sustainable evaluation practice (p. 444). Sustainable evaluation practice, or culture of evaluation, is described as organizational members who gather, analyze and use data and continually ask questions that matter to the ongoing improvement and growth of the organization (p. 444). Organizations tend to focus on involvement in the evaluation process, one-off training sessions, technical assistance and written materials as evaluation capacity building teaching and learning strategies, but they may not be appropriate or adequate depending on the context of the organization (Preskill and Boye, 2008b; Better Evaluation, 2015). There are multiple strategies that can be combined to be effective in a tailored evaluation capacity building design, including: technical assistance, written materials, technology, mentoring and coaching, meetings, communities of practice, internships and appreciative inquiry.

Mayne (2008) advocates for evaluative cultures within organizations, which is defined by organizations that: "engage in self-reflection and self-examination," "evidence-based learning," and "encourages experimentation and change" (p. 5). According to Mayne, the process of undertaking and conducting an evaluation provides evaluation capacity building opportunities for organizations, even more so than the learning that is presented in the form of a final evaluation report. Botcheva, White and Huffman (2002) echo Mayne's call for evaluative cultures and state that community-based organizations must view evaluation, learning and evidence-based practices as a core organizational function that builds organizations, both in terms of staff competency and competitive advantage for future funding. They also state that the existence of an evaluation or research department is a major catalyst for developing a culture of learning and evaluation.

El Hassar, Poth, Gokiert and Bulut (2020) emphasize the importance of organizational context when considering an evaluation capacity assessment tool. They developed a tool specifically for early childhood development nonprofits operating in Alberta, titled the Evaluation Capacity Needs Assessment. They argue that selecting an assessment tool that is not designed for the context in which evaluation capacity is being measured impacts the validity of the assessment results, and can have negative impacts on organizations, policies and people. Therefore, it is important for evaluators interested in assessing and building evaluation capacity in organizations to be cognizant that evaluation capacity assessment tools are often designed for a specific organizational context and not one size fits all.

The evaluation capacity building literature outlines a multitude of factors that correspond to evaluation capacity in organizations, tools to empirically measure evaluation capacity across identified factors, and activities and practices organizations can adopt to build evaluation capacity. The ultimate goal of evaluation capacity building interventions in organizations is to promote evaluative thinking and cultures of evaluation, which are a result of technical competency and resources to do evaluation, understanding and appreciation for evidence-based decision making at all levels of organizations, and the pursuit of evaluative knowledge not only as a means of accountability, but for organizational learning, reflection, growth and engagement. The central theme found in the evaluation capacity and capacity building literature is that organizations exhibit evaluative cultures when they facilitate and encourage collective enquiry, reflection, and the collection of data that is meaningful to all interested parties. One difference is that some authors emphasize the impact of external factors on evaluation capacity (Bourgeois and Cousins 2008; Volkov and King, 2007), where it is omitted by others (Preskill and Boyle, 2008a; Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry and Balcazar, 2013). External factors are significant in the context of this study because ISS is a nonprofit that heavily depends on public funding, which fluctuates based on public support for programming and political priorities.

2.5 Evaluation Practices and Capacity of Nonprofits

According to the literature, nonprofits have not kept pace with the ever-increasing demands for accountability from public funders, particularly in the field of evaluation and evidence-based practice (Bach-Mortensen and Mongomery, 2018; Botcheva, White and Huffman, 2002; Carman, 2007; Carman and Fredericks, 2010; Despard, 2016; Gill, 2010; Stoecker, 2007). Nonprofits face distinct challenges to evaluate their services, including lack of resources, technical capacity and skill for evaluation, ability to identify appropriate outcomes and indicators, organizational culture and resistance to evaluation, and lack of evaluation support and direction from funders (Bach-Mortensen and Mongomery, 2018; Botcheva, White and Huffman, 2002). In addition, nonprofits use various methods to collect evidence about their programs, but do not use that evidence to formally evaluate their programs (Carman, 2007). As discussed prior, evaluation is a systematic process to assess specific aspects of programs, policies and interventions. Nonprofits tend to equate data collection activities, such as surveys and focus groups, with evaluation without conducting the systematic and in-depth analysis of data that an evaluation entails (Despard, 2016). Furthermore, nonprofits face ever-increasing competition as the nonprofit sector continues to grow and information sharing and advocacy of social issues and services continues to increase in the general society, resulting in increased critique of the sector (Gill, 2010).

Carman (2007) researched the evaluation practices of nonprofits in the United States, and she found that nonprofit managers tend to equate evaluation and performance measurement with a host of accountability functions, including reporting, accounting, monitoring and regulation, and general management, but do little in terms of actual program evaluation which would inform both the outcomes of their services and tangible means to improve. Nonprofits collect data on many points of service delivery but struggle to collect outcome data, which can be attributed to lack of knowledge or skill of how to measure social outcomes and lack of support or direction from funders to do so. Furthermore, very few nonprofits surveyed had internal evaluation staff and instead delegated evaluation to frontline staff, support staff and managers who lack technical evaluation knowledge and skill. As a result, many nonprofits have yet to invest in evaluation as a key organizational function (Carman, 2007). Lastly, nonprofits tend to equate evaluation with accountability, to the detriment of organizational learning and long term systems change (Gill, 2010).

Mitchell and Berlan (2018) empirically analyzed the relationship between aspects of evaluation capacity in nonprofits, including evaluative culture, external and internal pressure, evaluative rigor, and monitoring frequency, and various characteristics of organizations, including size, age, budget surplus, staff compensation, and executive compensation. They found that larger nonprofits and those with budget surplus are positively associated with evaluation capacity and rigor, likely due to both increased pressure to evaluate from various sources and the available funds to pursue evaluation. Additionally, staff compensation is positively associated with evaluative culture and frequency as higher pay attracts candidates with higher education and technical credentials. Conversely, executive compensation actually has a negative effect on evaluative culture and practices, which they hypothesize could be due to the fact that higher paid and experienced executives may rely more heavily on their own judgement and experience rather than evidence-based practice.

Bach-Mortensen and Montgomery (2018) completed a systematic review of literature that address the barriers and facilitators of third sector organizations (TSOs) to evaluate their services. The most cited barriers to evaluation includes, in order of frequency: the lack of technical knowledge and ability to conduct evaluations, a perceived misalignment between the outcomes imposed by the funder and those prioritized by the TSO, and the lack of financial resources to conduct evaluations, either through contracting evaluators or having paid evaluation staff. The most cited facilitators to evaluation in the systematic review include: funders requiring TSOs to complete evaluations; the involvement of stakeholders in the identification of program outcomes and goals; and, having staff who have the technical competency and training to perform evaluations.

Stoecker (2007) surveyed nonprofits in the United States regarding their research practices, capacity and needs. The study found that many nonprofits devote enormous amounts of paid staff and volunteer hours to data collection and database entry for data that is seldom accessed or used. Furthermore, data is not well shared among organizations, which further limits the use and impact of the data collected. Stoecker further concludes that nonprofits appear to collect data for the purpose of performative accountability, resulting in large amounts of data that is seldom used for any purpose, let alone for tracking and measuring program outcomes.

Liket, Rey-Garcia and Maas (2014) write that nonprofits face a number of barriers to evaluate their programs, including: confusion about the type and quantity of data to collect; unclear messages from funders about preferred evaluation methods and evaluative rigor expected; and, lack of budget to contract evaluators or to have evaluators as regular staff, which then results in the offloading of evaluation to support staff and managers who lack evaluation expertise (p. 172). They argue that it is critical that nonprofits develop their negotiation skills with funders and other stakeholders around the topic of funding evaluation, so that nonprofits can more effectively meet the evaluation demands of funders while also producing evaluations that promote improvement and achievement of their social mission (Liket, Rey-Garcia and Maas, 2014).

Despite the close relationship that governments and nonprofits share, nonprofits have struggled to prioritize evaluation as a key organizational function. Rather, nonprofits view evaluation as an additional administrative burden as it relates to upward funder accountability, instead of maximizing the multifaceted benefits of evaluation, including developing and engaging staff, fostering organizational learning and further achieving their social mission. In addition, nonprofits struggle to differentiate between data collect methods, such inputting data in a database or distributing a paper survey, and the formal process of evaluation, such as the use of various lines of evidence to determine if a program has made an impact in the lives of participants. Nonprofits face a multitude of barriers to evaluation and to establishing evaluative cultures, including lack of technical skill and competency, resource deficits and volatile funding climates, difficulty in gathering data that accurately measures social outcomes, and working in silos due to competition for funding and the urgent nature of the nonprofit sector.

2.7 Gaps in the Literature

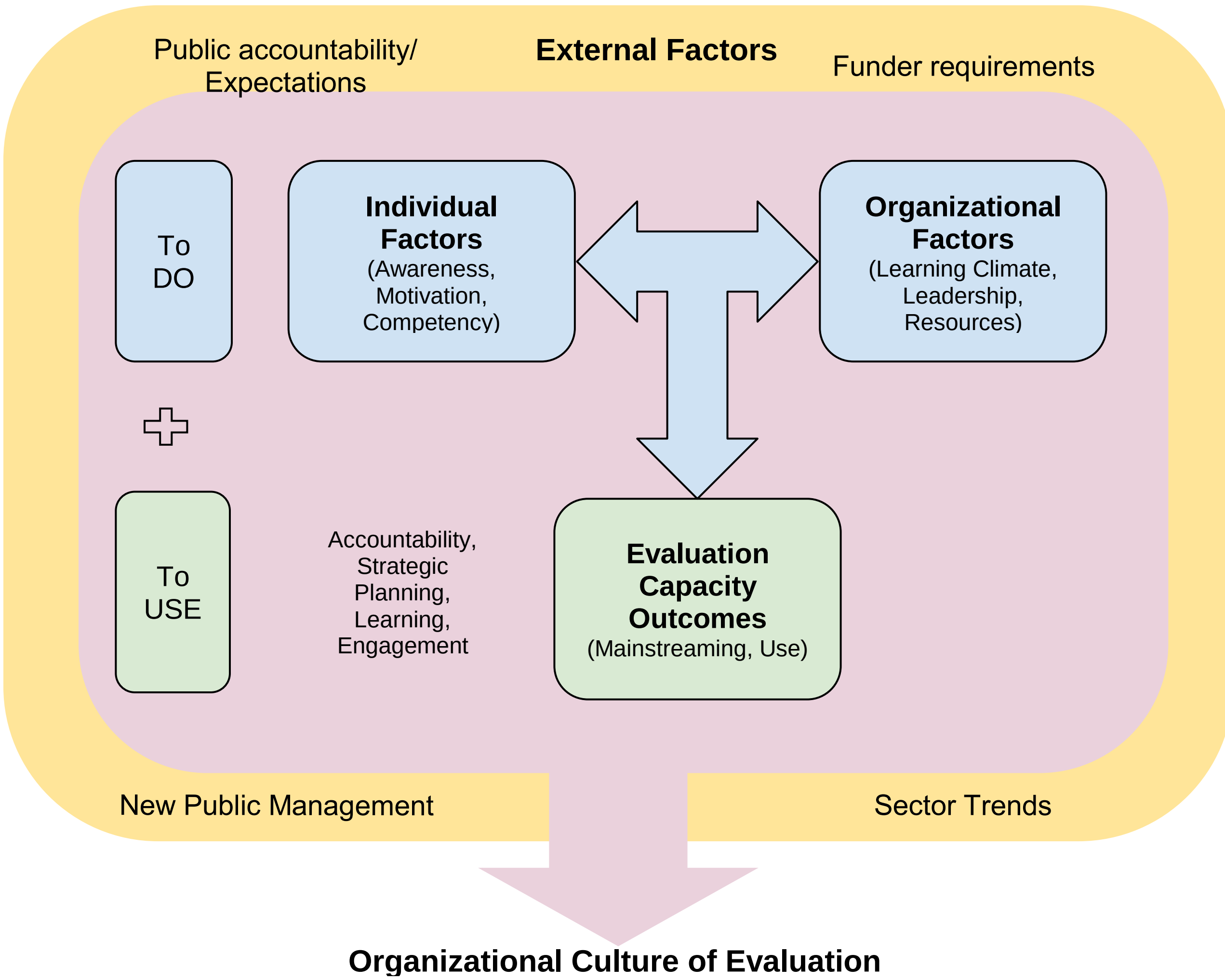
Despite acknowledgement in the literature (Preskill, 2014), there continues to be a lack of empirical studies of evaluation capacity building in organizations, particularly in relation to nonprofit organizations. There are multiple tools that can be used to measure evaluation capacity and instances where evaluation capacity has been measured, including recommendations for growth. However, there are few studies that measure the change in nonprofit organizational capacity, or rather, very few evaluations of evaluation capacity building interventions in organizations, which has been noted by Preskill (2014). It is worth exploring how nonprofits have assessed their evaluation capacity, implemented evaluation capacity building strategies based on empirical assessment, and then later reassessed capacity post-implementation of evaluation capacity building strategies to determine growth (or lack thereof) of capacity. This research would shed light on how nonprofits, despite the common barriers they experience, are able to build evaluation capacity, and thereby positively impact evidence-based practice, strategic planning, funding sustainability and overall ability to socially innovate.

Furthermore, there are few studies that address barriers to evaluation capacity that deal with the complexity of program participants, apart from typical nonprofit organizational challenges, such as lack of resources. Immigrant serving nonprofits, such as the organizational focus of this research, may face distinct barriers to evaluating their services directly with clients, including, but not limited to: clients that do not speak an official language, impacting the ability to gather qualitative data in focus groups and interviews; clients who do not read or write in an official language and/or lack literacy skills, rendering any written evaluation methods inaccessible; and, the cultural context of clients in which traditional evaluation approaches many not be familiar or valid. Culturally responsive approaches to evaluation, as outlined by Chouinard and Cram (2020), attempt to address such barriers to evaluation and are worth exploring further in the context of immigrant serving agencies in Canada.

2.8 Conceptual Framework

The following conceptual framework represents the major findings of the literature review and provides a roadmap for how the research findings will be understood and approached. Evaluation capacity is the organization's ability to do evaluation, which is an interplay between the individuals who make up the organization and the organization's structures and resources, and the ability to use evaluation, or the outcomes of evaluation capacity. The ability of organizations to do and to use evaluation, as a key organizational function and for multiple uses, such as accountability and learning, contributes to an overall culture of evaluation. Furthermore, multiple evaluation capacity building strategies or interventions can be used by organizations to develop individual, organizational, and evaluation capacity outcome factors. A visualization of the conceptual framework is included below in Figure 1.

Figure 1. Conceptual Framework of Evaluation Capacity



3.0 Methodology and Methods

This chapter provides an overview of the research methodology and design, the methods used to collect data, how data will be analyzed and project limitations.

3.1 Methodology

The project is based on a mixed methods case study design (Creswell and Plano Clark, 2017). According to Yin (2009), a case study is a methodological approach to research in which the researcher investigates a specific "phenomenon in depth within its real-life context" (p. 18). This project is a case study because it investigates the evaluation perceptions and practices of ISS within their real-life context, without using baseline data, a comparison group or a normative measure of an ideal organization (McDavid, Huse & Hawthorn 2017). The case study approach is to better understand how evaluation as a phenomenon, including data collection practices, staff experience and perception, and evaluation uses, contribute to a culture of evaluation in the organization. Furthermore, while the literature review has established factors that contribute to a culture of evaluation, the research is being approached from an exploratory perspective in that there is no ideal state the research attempts to measure the organization against.

The case study uses a mixed method design in which quantitative data is gathered first, followed by a qualitative method that builds on the analysis and findings of the quantitative data. This design is described by Creswell and Plano Clark (2017) as a sequential explanatory design (p. 77). According to Ivankova, Creswell and Stick (2006), the benefits of using mixed method designs is that neither quantitative nor qualitative methods are able to capture the nuances of a given issue; therefore, they can complement, enhance and fill in the gaps of either method.

The mixed method case study can be annotated using the research design depicted in Table 1. 'X' represents ISS as the organization exists and operates in its real life context, including its evaluation practices, perceptions and uses. Both 'O's represent the explanatory sequential data collection methods used to study the phenomenon or research focus, namely a quantitative survey followed by qualitative discussions.

Table 1. Mixed Method Case Study

X	O	O
Evaluation at ISS	Survey (Quantitative)	Small Group Discussions (Qualitative)

3.2 Methods

For this research project, a survey was distributed to all staff of ISS to gather quantitative data, which was followed by small group discussions with staff to present the survey findings and further discuss the evaluation practices of the agency and explore opportunities to build evaluation capacity as a group. The survey provides an opportunity to quantitatively measure evaluation capacity across the three domains of individual, organizational and evaluation capacity outcome factors (more detailed description below), while the small group discussions allow the topic of evaluation capacity, and specifically current evaluation practices, opportunities and challenges to build evaluation capacity, to be explored in more depth. The strength in completing the survey first is that specific areas of evaluation capacity, such as leadership and learning climate, can be highlighted, rather than discussing the topic in terms that are more abstract with participants. Furthermore, through the survey, staff are presented with the various factors related to evaluation capacity and they then have the opportunity to discuss these factors in more depth during the small group discussion, which is a capacity building activity in itself.

3.2.1 Evaluation Capacity Assessment Instrument

The survey distributed to staff is an adaptation of the Evaluation Capacity Assessment Instrument (ECAI), found in Appendix 1, which is an evaluation assessment tool specifically for nonprofit organizations (Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry and Balcazar, 2013). The tool was developed and validated by Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry and Balcazar who hold positions at American and Irish universities. The original tool has 68 items that address three evaluation dimensions: 1) Individual Factors, including awareness, motivation and competence; 2) Organizational Factors, such as leadership, learning climate and organizational resources; and 3) Evaluation Capacity Outcomes, namely integrating evaluation practices into everyday activities. The ECAI has been used to assess evaluation capacity by Lindeman, Bettin, Beach, Adames, Johnson, Kern, Stonehouse, Green and Phillips (2018), Stitt-Bergh (2016) and Suarez-Balcazar, Taylor-Ritzler and Morales-Curtin (2015). For this project, the survey was adapted to better fit the context of ISS. The original tool has 68 statements, but five questions were removed that were not considered relevant to the context of ISS. Furthermore, the wording in some of the statements was adapted to the language of the organization, such as using the word “service” and “agency” rather than “organization”.

Each ECAI statement is rated on a scale of 1 to 4: 1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, and 4 = strongly agree. Hereafter, all references of scores or ratings of statements should all be understood as a score or rating out of 4. Scores closest to 1 signify lower capacity whereas scores closest to 4 signify higher capacity. However, it should be noted that the ECAI is an exploratory tool and not a diagnostic tool, therefore the scores are meant to draw attention to areas of strength and growth, rather than specifically diagnose an organization of having low or high capacity.

In a review of the literature, there were other validated evaluation capacity measurement tools found (Bourgeois and Cousins, 2008; Bourgeois, Whynot and Theriault, 2015), which are developed in the context of government agencies, and specific sub-sectors of social services, such as early childhood development (El Hassar, Poth, Gokiert and Bulut, 2020). However, the ECAI is the only validated instrument found that is general to nonprofit organizations. Therefore, it most adequately fits the context of ISS compared to other tools. The organizational focus of the tool and context of the study are important to align, given that government agencies and nonprofits have significant differences, including resources available, evaluation context, and time available to staff to pursue evaluation, among others.

In August 2020, I announced the project at the ISS All-Staff meeting, which was attended remotely by 77 staff members. As noted previously, I have worked at ISS for the past five years, so the announcement was made to my colleagues and peers. The announcement described the purpose of the study, informed staff that a survey would be emailed to all staff in September, and the nature of questions they can anticipate in the survey. To encourage participation, I described the survey in terms of organizational culture and how ISS gathers and uses information, given that many staff may perceive evaluation as a laborious and technical practice that does not relate to their day-to-day work.

The ECAI was input into Survey Monkey, the digital survey platform available to University of Victoria students and researchers. On September 10th, the initial invitation to participate and survey link was emailed to the three listservs of ISS, including the email list for Edmonton East Office, Edmonton Reception House, and Red Deer staff. These lists included all staff working at the three locations who have an agency email address. Relief staff, such as Community Support Workers, often do not have agency email addresses and therefore were not recruited for the study. The initial invitation described the purpose of the survey, benefits to participation, time required, and risks to participation. Following the first email invitation, a reminder email was sent out to all-staff once a week, each decreasing in formality. The survey was closed on October 6th after a total of four emails were sent to staff in the three aforementioned listservs.

The survey had 41 respondents out of 115 regular full-time and part-time staff, resulting in a 36% response rate. The total breakdown of respondents is found below in Table 2 (Position of Respondents and Actual Staff Count) and Table 3 (Tenure of Respondents). No other demographic information was gathered regarding the respondents. Given the size of the organization, collecting additional demographic data related to the participants could decrease my ability to protect participants' anonymity. As noted in Table 2, the respondents represent a cross-section of the service. Programs Managers and Team Leaders had high participation relative to their actual presence in the service, whereas Settlement Counsellors had the lowest participation. This difference can likely be attributed to managers' understanding of evaluation and its significance, whereas frontline staff have busy schedules due to dealing directly with clients day-to-day. Furthermore, lack of participation by frontline staff could also contribute to the perception that evaluation has less to do with their day-to-day functions, compared to management functions. The survey also had a 15% dropout rate; the demographic and individual factors sections had 41 participants, whereas the evaluation capacity outcomes section had between 36 and 35 respondents per question.

Table 2. Position of Respondents and Actual Staff Count

Position	Respondent Count	Respondent Percentage	Actual Staff Count	Actual Staff Percentage
Program Manager / Senior Leadership	5	12%	9	8%
Team Leader	9	22%	10	9%
Program / Project / Volunteer Coordinator	7	17%	17	15%
Settlement Counsellor	8	20%	40	35%
Language Assessor	3	7%	11	10%
Program Assistant	2	5%	3	3%
Other Support Staff	7	17%	25	22%
Total	**41**	**100%**	**115**	**100%**

Table 3. Tenure with Agency of Respondents

Tenure with Agency	Respondent Count	Respondent Percentage
Less than 1 year	0	0%

2 - 5 years	16	39%
5 - 10 years	9	22%
10 - 15 years	13	31%
More than 16 years	3	7%
Total	**41**	**100%**

3.2.2 Small Group Discussions

After the survey concluded, a selection of staff were invited by email to participate in small group discussions with their peers. I emailed all of the participants privately and described the purpose of the small group discussions, time commitment, and format. The email also included the participant consent form to ensure that participants were aware of the voluntary nature of the study, how their confidentiality would be protected, and potential benefits and risks of the study.

The sample can be categorized as an expert or purposeful sample, because all of the staff invited either have experience evaluating their own programs or are known by the researcher to have attended evaluation training or capacity building opportunities (Frey, 2018, Patton, 1999). Furthermore, the selection of participants was influenced by the survey results as Program Managers, Team Leaders and Program Coordinators all had the highest representation in the survey relative to their actual presence in the agency. The small group discussions were intended to better understand staff perceptions of individual, organizational and mainstreaming factors related to evaluation, so it was imperative to invite staff who had some understanding of evaluation concepts to facilitate discussion.

The small group discussions were organized into three groups: Coordinators, Team Leaders and Program Managers. The groups were organized based on position to ensure participants were not participating in the discussions with their direct supervisor in an effort to reduce power differentials between participants. Thirteen staff agreed to participate in the small group discussions remotely on Microsoft Teams during the last week of October. The breakdown of participants in the small group discussions can be viewed in Table 4. The following programs were represented at least once in the three discussions: Intercultural Education, Family Wellness and Community Enhancement Program, Orientation and LINC Outreach, Volunteer Programs, Language Assessment, Newcomer Settlement Program, Community Support Services Program, Refugee Health Promotion and the Resettlement Assistance Program. Some participants represented multiple programs, either by currently working in the program or during their tenure with the Service. Similar to the survey, no other demographic information related to the participants will be shared in an effort to protect the confidential nature of the study.

Table 4. Position of Small Group Discussion Participants

Position of Participants	Count
Program Manager	3
Team Leader	5
Program / Project / Volunteer Coordinator	5[1]
Total	**13**

The small group discussions followed this agenda:

1. Review of the consent form, including the voluntary and confidential nature of the study;
2. Introduction of the principal researcher's interest in pursuing the research and the purpose of the study;
3. Definition of evaluation, as an organizational practice and cultural orientation;
4. Introduction of participants, their experience and knowledge of evaluation, and how their program is evaluated;
5. Description and purpose of the ECAI;
6. Presentation of the ECAI results at ISS; and,
7. Facilitated group discussion.

Items number six and seven in the list above happened simultaneously in that the ECAI findings were presented in stages. Participants were presented the survey results, grouped according to three dimensions that structured the survey: 1) Individual Factors, including awareness, motivation and competence; 2) Organizational Factors, such as leadership, learning climate and organizational resources; and 3) Evaluation Capacity Outcomes, namely integrating evaluation practices into everyday activities. I also explained the scale of the survey and how the scores were averaged to create an overall score out of four for each statement. A discussion with participants took place following the presentation of results for each subfactor. At this point, participants were asked if the results were similar to their experience, if the results surprised them, and if they had any thoughts related to the results (Sample Slide in Appendix 3). The process was repeated for each sub-factor. Once the survey results concluded, the discussion shifted to what ISS does well in terms of evaluation, what could be improved, activities that could grow evaluation capacity, and how data that is already collected could be better used for evaluation purposes, including outcome measurement, strategic planning, and organizational learning. All three of the small group discussions lasted two hours.

Each of the discussions were video recorded on Microsoft Teams and transcribed. Participant confidentiality was protected starting at the transcription process, as participant names,

[1] One Coordinator participated in the group discussion with Team Leaders due to scheduling availability, after it was determined that their direct supervisor would not be present in the discussion.

colleague's names referenced by the participants, and program names, were all omitted from the original transcription.

It is important to note that I am a colleague of the discussion participants. I actively participated in the discussions, particularly when the participants referred to my role in supporting the evaluation of their programs, such as conducting focus groups with their clients. As well, throughout the small group discussions I emphasized that the ECAI is an exploratory measurement tool. Explicitly, it is not intended to diagnose a specific evaluation capacity building intervention, but encourage the participants to reflect on their evaluation practices and perceptions and consider areas that they could build evaluation capacity. As such, I was situated in the discussions as a colleague and collaborator, rather than objective assessor of my colleagues' practices and expertise.

3.3 Data Analysis

The survey results were analyzed first prior to conducting the small group discussion, following the logic of the sequential explanatory design (Creswell and Plano Clark, 2017). The survey results were shared during the group interviews to provide a starting point and catalyst for the discussions. The small group discussions provided an opportunity to explore and build on the key findings from the survey.

3.3.1 Survey Analysis

The survey data was compiled and organized to show the mean and mode for each survey statement, focusing on the highest and lowest scoring statements within each sub-factor. Bar charts were created for each sub-factor to demonstrate the differences in scores for each statement within each sub-factor. The highest and lowest scoring items highlighted key themes within the data that was then presented in the small group discussions and used as a starting point in the conversation about evaluation capacity within the organization.

3.3.2 Small Discussion Group Analysis

The three small group discussions were recorded and transcribed. Each of the three transcriptions were read three times; first, a preliminary reading to become familiar with the data; second, they were read closely and notes were taken about significant points or highlights in the data; and lastly, they were read again to better understand how the data answered the research questions. The data was then arranged to address what the evaluation capacity of ISS is, including our ability to do and to use evaluation, and facilitators and barriers to evaluation capacity, which directly responds to the research objectives.

3.3.3 Integrating Results

The findings of both the ECAI survey and the small group discussions were then arranged in a table to better understand how both data sets either complimented or challenged one another. The idea for the table is taken from Creswell and Plano Clark (2017, p. 239) as a means to demonstrate the relationship between the two methods.

3.3.4 Conceptual Framework

Lastly, the findings were then arranged into the conceptual framework graphic in the discussion and analysis section. The graphic provides a visual interpretation of the findings to enhance the accessibility of the findings for readers from the organization. The visual also provides a clearer interpretation of how the findings from the two methods could then bridge into recommendations for evaluation capacity building interventions for the organization.

3.4 Ethical approval

The project was reviewed and determined to have minimal risk of harm to participants. The project proposal was submitted to the Human Research Ethics Board at the University of Victoria in the spring of 2020 and was formally approved on May 29, 2020.

3.5 Research credibility and generalizability of results

Credibility refers to the techniques and designs used by qualitative researchers to ensure the trustworthiness of their findings and in their capacity as researchers in absence of experimental designs (Lincoln and Guba, 1985). Throughout the research process, I adopted various strategies to ensure high quality and credibility of the results. First, I used a validated evaluation capacity measurement tool for non-profit organizational contexts. The use of a validated tool implies that the use of the individual factors, organizational factors and evaluation capacity outcomes, and the corresponding statements attached to each factor in the survey were validated and found to be accurate measures of each construct.

Lincoln and Guba (1985) note a number of techniques for promoting the credibility of qualitative findings, including: prolonged engagement, triangulation and member checking. They describe prolonged engagement as the commitment and investment of the researcher in learning the culture of an organization, its background and objectives, and building trust with members. I practiced prolonged engagement with the organization by nature of my employment at ISS over the past five years. I am an insider of the organization and therefore have lived experience and knowledge with the organization and in the sector, which provided a nuanced understanding of how evaluation is carried out and used in the organization and the distinct challenges the organization faces. I was also able to readily build trust and rapport with the small group discussion participants because I was among colleagues who I have worked with over the past five years. Furthermore, being an insider of the organization provided the opportunity to continually debrief the project and its findings with my colleagues, which Patton (2009) refers to as perspective triangulation and Lincoln and Guba (1985) refer to as peer debriefings.

Triangulation refers to the use of numerous lines of evidence, data collection methods, perspectives and theories to answer the research question (Lincoln and Guba, 1985). In this project, the topic of evaluation capacity within the organization was explored in depth through a survey and small group discussions and then integrated in the analysis to understand how the findings from both methods either complimented or challenged each other, which provided an opportunity to triangulate different lines of evidence and methods. The use of both distinct methods compensates for the weaknesses present in either the survey or group discussions had they taken place in isolation. Throughout the research project, I also undertook perspective triangulation (Patton, 2009) with colleagues, including conversations with colleagues over the past two years regarding the evaluation practices of ISS, during the literature review to assess whether certain topics were salient to the research, and again in the data analysis. Furthermore, in my current role I often attend evaluation webinars and learning opportunities specific to the newcomer serving sector in Alberta and across Canada. These learning opportunities have served as a less formal means to triangulate findings; the results of the literature review and the conceptual framework produced have been confirmed as valid constructs of organizational evaluation capacity in these supplementary learning opportunities.

The third notable approach to establish credibility is through member checks, which are described by Lincoln and Guba (1985) as the process in which key stakeholders, for whom the study was undertaken, review the findings and assess their utility and applicability. I engaged with the client for this research throughout the project, including in debriefs about the literature review, data collection, and most recently to review the draft of this report. The client was able to review the drafted report in its entirety, including the findings and recommendations, and confirm their applicability and accuracy.

Although my insider status in the organization is a strength, it also poses a limitation to the research. I had to ensure my colleagues understood the confidential and voluntary nature of the study and that their participation in the small group discussions were taking place within the context of academic research and not an informal conversation with their peers. As such, each participant signed a consent form that indicates how the discussion data will be used, including the precautions that will be taken to ensure each participant's responses are confidential and that their anonymity is protected within the final report. Regardless, participants could have felt uncomfortable sharing their personal opinions about the limitations of our evaluation processes and speaking about evaluation capacity from a perspective of deficit in front of myself and their peers.

An additional limitation is that the survey had a 36% response rate. A similar study carried out by Fierro and Christie (2017), which looked at manager perceptions of evaluation in American grantee programs of federally funded health agencies, had a 40% response rate. Despite having respondents across positions and tenure with the agency, there was a low response from frontline staff. The research process formally began during the COVID-19 pandemic. On March 17, 2020, ISS shifted almost all of their services to remote delivery, which resulted in staff working from home and using new technology platforms, such as Office 365 and Microsoft Teams. Most staff have experienced a significant learning curve to using these new technologies, and have informally expressed being overwhelmed with the new technological aspects of their work, which subsequently could have negatively impacted the survey response rate as it was sent as a digital survey by staff email in September 2020 while many staff were still in the process of adjusting to the new work-from-home context.

Another factor that impacts the study is that evaluation can be perceived as a technical process to which frontline staff have little connection or formal training in. Therefore, it is possible that many frontline staff self-selected not to participate in the survey, despite my emphasis on the organizational culture aspects of evaluation. Furthermore, staff who feel more comfortable with the topic of evaluation may be over represented in the survey. Both of these factors could impact the validity of the survey findings.

In addition, a considerable factor related to evaluation capacity is the technical capacity to do evaluation, financial resources to complete evaluation, staff having the time to dedicate to evaluative activities, and physical documents used in evaluation, such as program logic models and evaluation methods, including paper surveys, among others. This information could be gathered by looking at the educational profiles of staff in Human Resource documents, actual budget to complete evaluation, the workload of staff, such as the actual time contributed to data collection and evaluative activities, and analysis of program documents related to evaluation. However, this study specifically explores staff perceptions, opinions and experiences with evaluation, through both the Evaluation Capacity Assessment Instrument and the small group discussions. These additional aspects of evaluation capacity within the organization could be explored in further study, which would enrich the context of the study.

Lastly, this research is carried out within a specific context in an organization. The project has not sought to produce generalizable findings across all nonprofits, but rather to understand evaluation capacity within ISS as a specific context. As stated by Lincoln and Guba (1985), the responsibility of the researcher lies with making their research context known, but it is the responsibility of the report user to determine whether the findings are applicable to their context. As a result, the findings and recommendations from this project may provide insight to similar organizations, such as other IRCC-funded immigrant serving organizations of similar size and overall capacity, but should not be approached as generalizable findings across varied organizations. Elements such as the conceptual framework and the data version in section 5.1 may be elements that could be used in different contexts when replicating such a study.

4.0 Findings

4.1 Introduction

The survey data was exported from Survey Monkey and organized into a table to group the ratings for each factor, including the average and mode for each statement, as seen in Appendix 2.

4.2 Evaluation Capacity Assessment Instrument

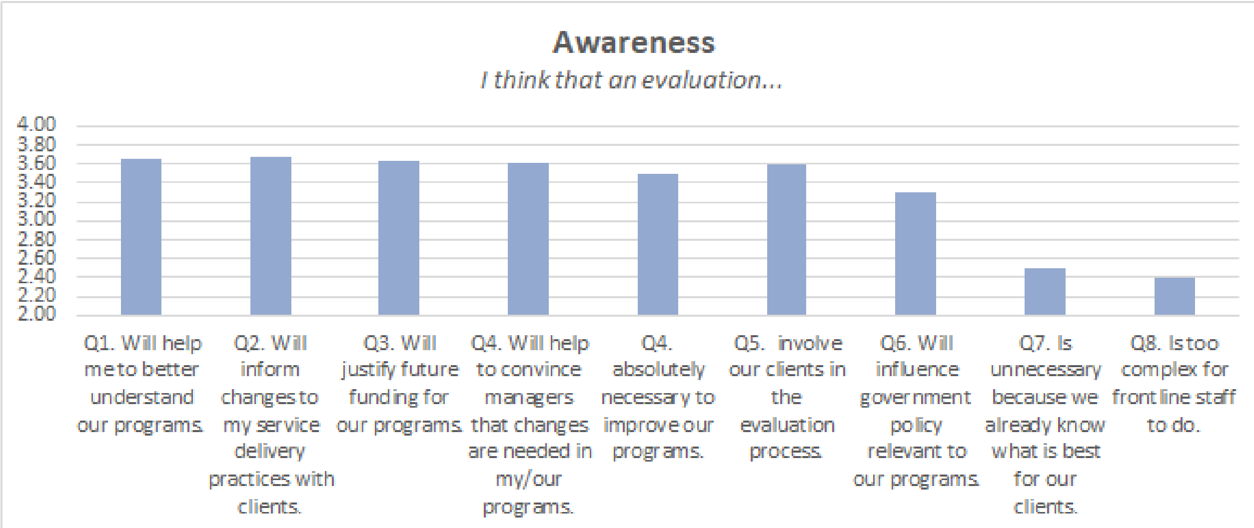
4.2.1 Individual Factors

The individual factors include the sub-factors of awareness, motivation and competency. Overall, motivation and awareness had higher response averages than competency. Motivation scored 3.59, awareness scored 3.31, followed by competency at 3.06. The motivation statements also had the least variance between statement scores compared to awareness and competency, although there were only four motivation statements compared to 12 competency statements. Motivation statements have a variance of 0.21 between the highest and lowest rated statements whereas awareness had a variance of 1.29 and competency of 1.34. The variance in scores between the individual sub-factors could support that staff have less understanding of the varied uses and applications of evaluation practice and also have varied competencies for evaluation, but are interested and motivated to learn more about it.

Participants rated the following awareness statements the highest: Q1) awareness that evaluation will help to better understand programs (3.66) and Q2) will inform changes to service delivery practices (3.68). On the contrary, participants rated the following awareness statements lower: Q7) perception that evaluation is unnecessary because staff already know what is best for their clients (2.49) and Q8) evaluation is too complex for frontline staff to do (2.39)[2]. These results could indicate that staff perceive evaluation as being important to their work but is too difficult or labour intensive for themselves to do.

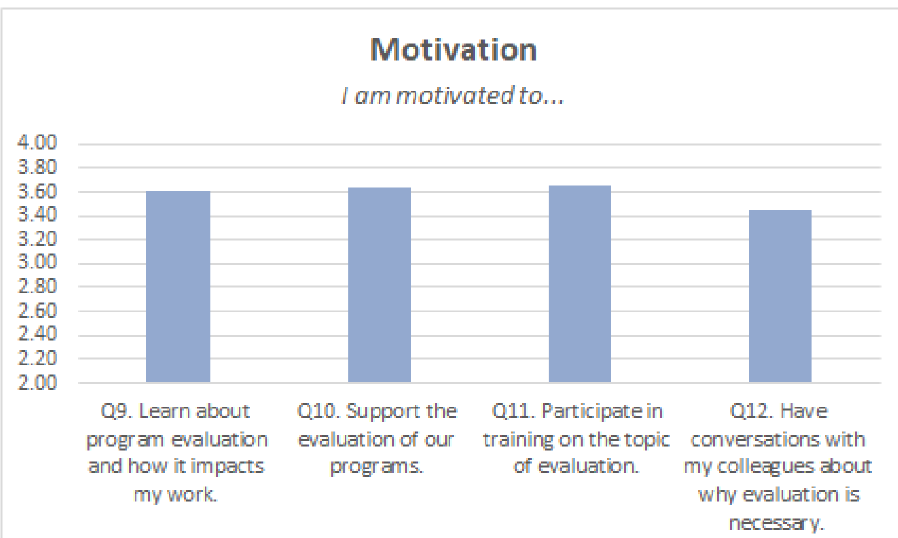
Figure 2. Bar Chart of Awareness Responses (n=41)

[2] All negative statements had their scores inverted in the data analysis. Therefore, the actual score for Q7 "Evaluation is unnecessary because we already know what is because for our clients" was 1.51, but was inverted resulting in a score of 2.49 to reflect the negative wording of the statement. Although there were higher levels of disagreement with the statement, it still scores lowest relative to other statements within the sub-factor.



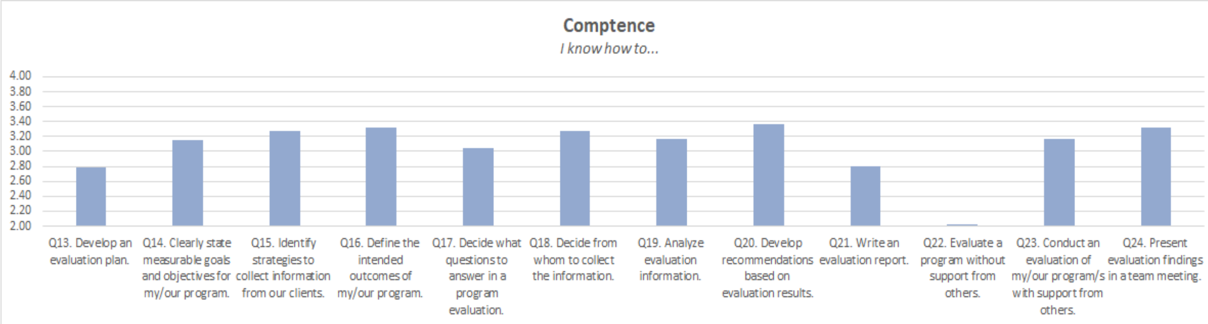
As noted previously, the motivation statements were rated highly by participants, with Q11) motivation to participate in evaluation training opportunities scoring the highest (3.66).

Figure 3. Bar Chart of Motivation Responses (n=41)



Finally, the competency statements had the most variance in average ratings. Participants rated the following competency statements the highest: Q16) knowledge how to define program outcomes (3.32) and Q20) knowledge how to develop recommendations based on evaluation results (3.37). They rated the following competency statements the lowest: Q13) ability to develop an evaluation plan (2.78) and (Q22) ability to evaluate a program without support from others (2.02).

Figure 4. Bar Chart of Competency Responses (n=41)



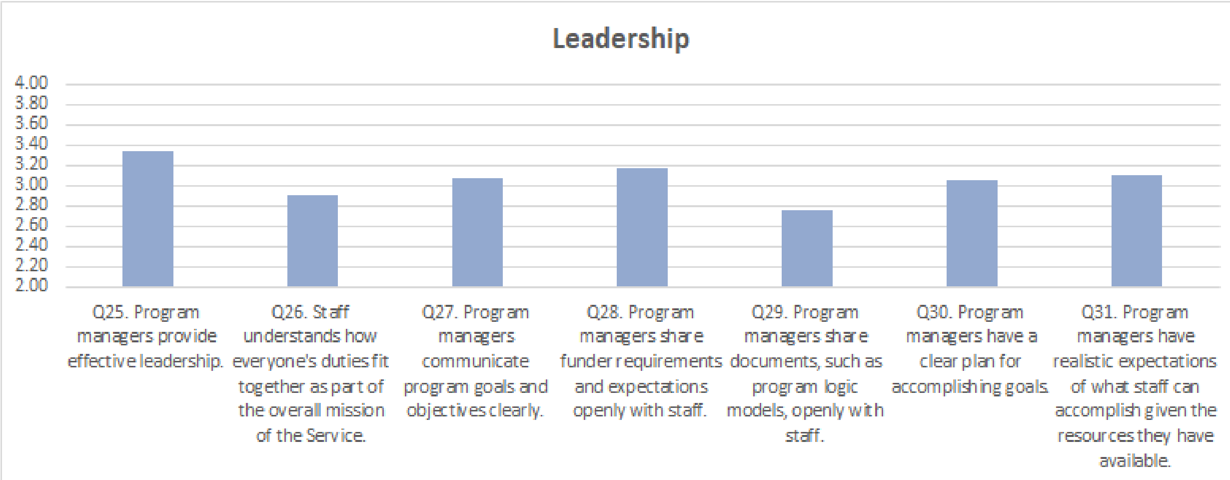
Overall, results from the Individual Factors section of the ECAI show that staff are aware of the benefits of evaluation, its potential positive impacts on the organization and their programs, and they are motivated to learn more about evaluation and support the evaluation of their programs. As the individual competency statements vary, there could be a benefit to focusing training opportunities on more technical aspects of evaluation, such as how to develop an evaluation plan for programs, determining appropriate methods and research questions to ask in an evaluation, and how to summarize evaluation findings in a report.

4.2.2 Organizational Factors

The organizational factors include the sub-factors of leadership, learning climate, and resources. All survey participants were asked to rate leadership and learning climate statements, whereas the resources statements were split into questions for all participants and questions for Team Leaders and Management only, given that some questions addressed financial resources and relationship with the funder. The leadership statements scored the highest with an average rating of 3.05, compared to learning climate at 2.91 and resources at 2.98. Leadership also had the most agreement between statements with a difference between the highest and lowest scoring statements of 0.58, whereas the learning climate had a variance between statements of 1.33.

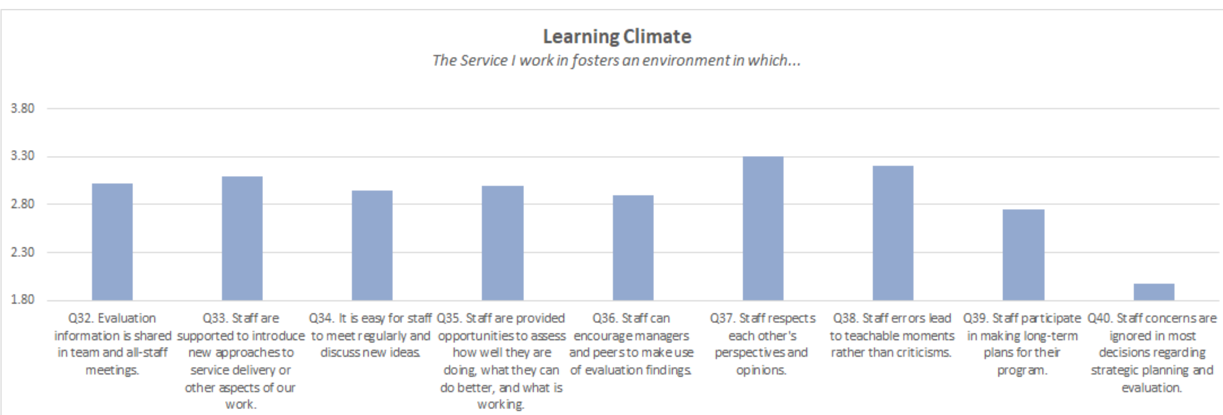
The highest rated leadership statements include: Q25) program managers provide effective leadership (3.33) and Q28) managers share funder expectations openly with staff (3.18). The lowest ranking leadership statements, relative to other statements in the sub-factors, include: Q26) staff understand how everyone's duties fit together with the overall mission of the service (2.90) and Q29) that managers share program documents, such as logic models, openly with staff (2.75). The sharing of program documents, including program and funder logic models, could be a straightforward intervention for staff not understanding how everyone's duties fit together with the overall mission of the service.

Figure 5. Bar Chart of Leadership Responses (n=41)



In terms of learning climate, staff rated Q37) perception that staff respect each other's perspectives and opinions (3.30) and Q38) that errors lead to teachable moments rather than criticisms (3.20) the highest within the learning climate sub-factor. The lowest rated statements are Q39) staff are included in long term decision making (2.75) and Q40) staff concerns are mostly ignored in strategic planning and evaluation (1.98). Again, due to the negative phrasing of Q40, the final average was inverted. Overall, the results indicate that there is a positive learning culture at ISS that does not penalize staff for errors, but that staff could be further included in setting the longer term vision and strategy for the service.

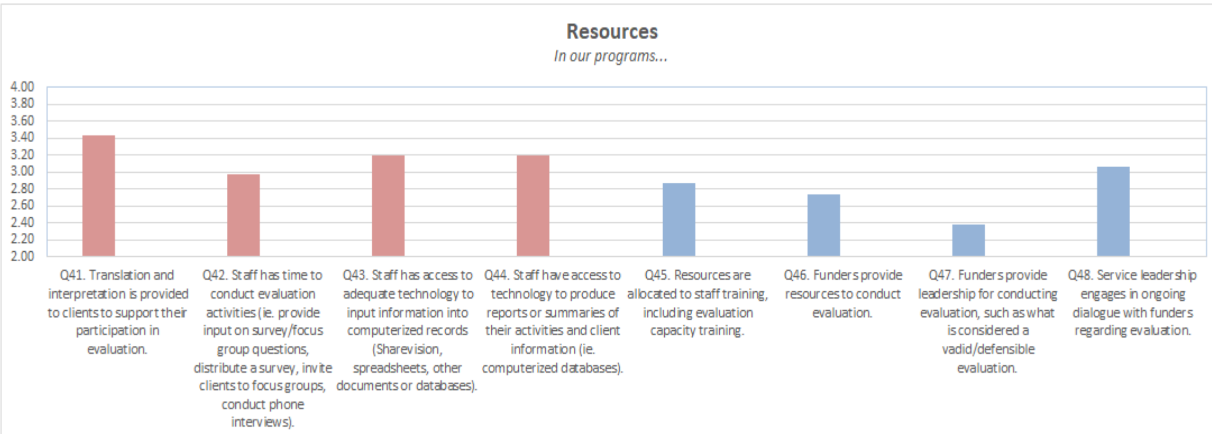
Figure 6. Bar Chart of Learning Climate Responses (n=40)



Lastly, the resources provided to evaluate programs had the highest variance between responses. The highest rated statements were Q41) that resources are provided to facilitate client participation in evaluation (3.43) and Q43 and Q44) staff have access to technology to input data, including computerized databases (3.20). However, team leaders and managers rated resource statements related to funders the lowest, including Q46) the funder provides resources to conduct evaluation (2.73) and Q47) the funder provides leadership for conducting evaluation (2.38). The Q42 and Q45 statements also scored below three, indicating lack of time to complete evaluation and resources to provide further evaluation training. Therefore, access to technology is not a big issue for the service's overall evaluation capacity, as was found in the literature as a common issue for nonprofits. However, the biggest resource constraints as evidenced in the survey are financial resources to complete evaluation, such as hiring an internal evaluator, evaluation leadership and guidance from the funder, and the time required to carry out evaluation.

Figure 7. Bar Chart of Resources

Red bars represent all staff (n=40). Blue bars represent Team Leader and Management responses (n=15).

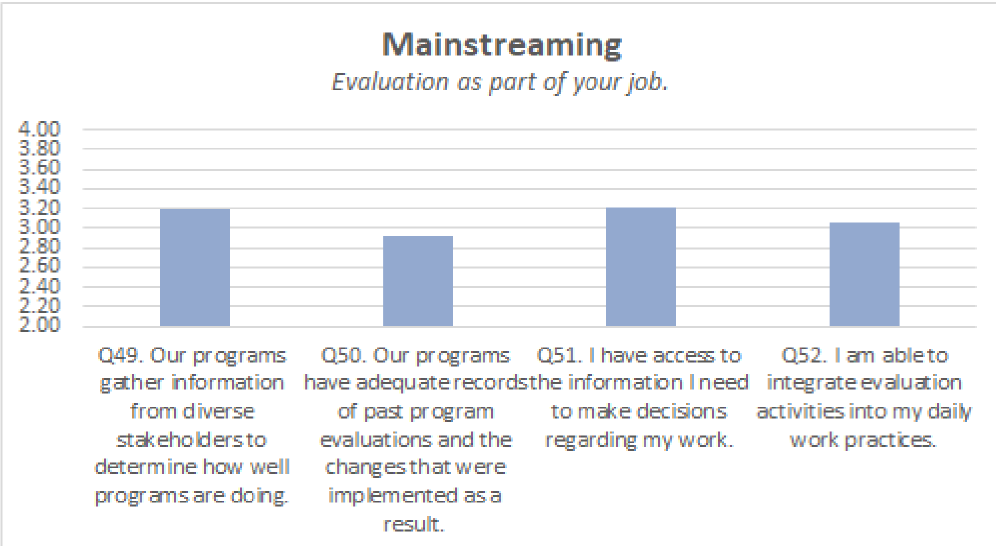


Taken together, the responses demonstrate an overall confidence in the leadership of ISS, although staff have less understanding of how their duties fit together into the broader picture of the service. This issue could be addressed by sharing higher level program information, such as logic models and expected outcomes of funded programs at the federal level, openly with staff through internal learning opportunities. The responses to learning climate and resources for evaluation had more variance in ratings between statements. Particularly, team leaders and managers rated statements related to funder guidance and resources for evaluation the lowest, which could then be connected to the overall learning climate. Given that the federal funder supports approximately 80% of all funding at ISS, it can be posited that their lack of guidance and provision of resources for evaluation heavily impacts the culture of evaluation at ISS and subsequent learning climate.

4.2.3 Evaluation Capacity Outcomes

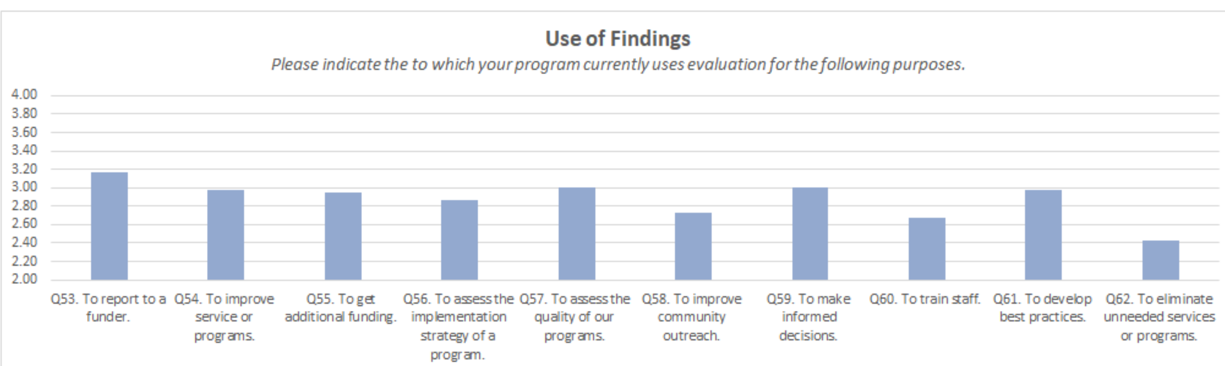
The Evaluation Capacity Outcomes factor includes organizational mainstreaming of evaluation and use of evaluation findings. The highest rated statements include: Q51) staff have access to information they need to make decisions regarding their work (3.22). The lowest rated include Q50) having records of past evaluations that staff have access to (2.92). Therefore, although staff may not have access to evaluative data regarding their programs, they feel informed enough to make decisions about their clients and programs, which complements the response for Q7 in that staff feel they may already know what is best for their clients regardless of evaluations taking place. There was a fairly positive response to Q52) staff having the ability to integrate evaluative activities into their daily work (3.06), which contrasts with the previous score for resources and staff having a lowest response to statements regarding time to complete evaluative activities.

Figure 8. Bar Chart of Mainstreaming (n=37)



There was greater variance in responses for uses for evaluation. Not surprisingly, respondents rated Q53) using evaluation to report to funders (3.17) the highest, followed by using evaluation to Q57) assess quality of programs (3.00) and Q59) make decisions (3.00). The lowest ranking uses were Q60) using evaluation to train staff (2.68) and Q62) to eliminate unneeded programs or services (2.43). Similar to the responses for leadership and resources, the findings indicate that the majority of evaluative activities at ISS are dependent on and driven by the funder, rather than internal learning.

Figure 9. Bar Chart of Use of Findings (n=36)



4.2.4 Open Ended Question

The survey concluded with an open-ended question that only garnered six responses, including: the need to increase resources to complete program evaluations, including for evaluation training and having an internal staff evaluator; communicating evaluative information more openly among staff and clients; and, the challenge of completing evaluations with clients.

4.3 Small Group Discussions

Three small group discussions were held with Coordinators, Team Leaders, and Program Managers. The findings from the discussions were separated into four key themes: how the organization is currently doing evaluation, how they are using evaluation, barriers to evaluation, and facilitators to evaluation. The differences between the perceptions of evaluation among the different groups is also notable and will also be discussed throughout the four subsections.

4.3.1 Doing Evaluation

The evaluative activities, or methods used, in ISS programs include surveys, pre/post surveys, and client follow-ups by phone and in-person, focus groups and interviews. Internal evaluation activities carried out by staff include paper and digital client surveys, client follow-ups by phone and in-person, and client focus groups. External evaluation activities carried out by contracted evaluators have included client surveys and individual interviews. In one instance described by a participant, a graduate student completed their thesis at the organization by evaluating an outdoor wellness program. They used a mix of qualitative methods, including interviews and photovoice. The final report was used for further program development, internal and external justification for the program and presentations at conferences. It was perceived by the participants that this evaluation was more meaningful because it collected first-hand accounts from clients of program impacts rather than "sterile" survey data.

Table 5. Inventory of Current Evaluation Methods

Program[3]	Method	Participants	Use of Data
Family Wellness and Community Enhancement Program (ended September 30, 2020)	Contracted third-party evaluator, post-evaluation (digital), one-on-one interviews	Clients	External accountability (funder), knowledge sharing with staff
Intercultural Education (Staff Training)	Post-evaluation (paper)	Staff	Internal accountability (West Office), program improvement
Intercultural Education (Parenting in 2 Cultures, Managing Anger, Raising Teenagers)	Pre/post evaluation (paper)	Clients	External accountability (funder), program improvement
Language Assessment, Referral and Counselling Centre (LARCC)	Post-evaluation, follow-ups (in-person or by phone), focus groups	Clients	Accountability (funder), program improvement, internal program meetings and updates
LINC Outreach	Post-evaluation (paper), thank-you cards from clients	Clients, LINC Teachers	Program improvement
Newcomer Settlement	Post-evaluation (paper and	Clients	Accountability (funder),

[3] The Resettlement Assistance Program was present in a small group discussion, but did not outline their current evaluation practice during the introduction so their current practices are not listed in the table.

Program	digital), bi-annual focus groups		program improvement
Orientation	Post-evaluation (paper), focus groups	Clients	External accountability (funder), program improvement
Volunteer Programs	Post-evaluation (paper), bi-annual focus groups, in process of developing digital pre/post evaluation	Clients, Volunteers	External accountability (funder), program improvement

One notable finding is that some participants named and explored various methods to collect data about their programs and tended to equate data collection methods with evaluation. However, these methods are used to collect lines of evidence about programs, but not to systematically test effectiveness of the programs in achieving their intended outcomes or to explore implementation strategies. It may be worthwhile to begin evaluation capacity building efforts on supporting staff to distinguish the difference between data collection methods and program evaluation. This process could be undertaken by: working with the staff of one program to look at a small sample of surveys that have already been collected; compiling the data into an easily analyzable format, such a spreadsheet; analyzing the data according to the established program outcome; and, reporting the findings at a team meeting.

4.3.2 Using Evaluation

Participants of all groups expressed that data gathered through various methods, such as surveys or focus groups, are typically used for internal and funder accountability in reports. One coordinator mentioned that there is often a push from management to distribute client surveys prior to a report being due, but that there is little follow-up or use of the survey data after the report has been submitted. Another participant mentioned that focus group data is used in narrative reports, stating that quotes from a focus group transcript will be extracted and placed into a report if it addresses a contract outcome, but there is little other use of the data. Two coordinators expressed that their clients regularly participate in focus groups, but that they do not know how the information is used. One coordinator expressed that they do receive the results of the focus groups conducted with their clients and that the information is used to improve their program. The others were surprised because they had never heard feedback from the focus groups and to their knowledge, no changes have been made to their programs as a result. Participant quotes that reference use include:

- “I think evaluations are used in a very informal way and only in the short term – the monthly report. I need a story right now, I need something right now, I need to evaluate right now! Let’s look at what happened last month and if there is something I can use. It’s not to make changes in the program long term.”
- “And where does the information go? In the focus groups.... I saved the decent quotes and put them in a report. Saved in the Ldrive, and that’s it. Even with my program, there were two questions that were used and the rest was fluff. If you get a decent quote, you put it on a brochure.”

At the Coordinator and Team Leader level, it was expressed that paper surveys were developed by individual staff within their program and used for program improvement purposes and better understanding client needs. In terms of group programs, Coordinators and Team Leaders said that they prefer evaluations that ask clients what topics they would like to learn more about, rather than asking for critical feedback because it can be culturally and linguistically challenging for clients to express. Group-based programs have taken it on themselves to distribute paper surveys for program improvement and finding new topics for group orientation sessions. However, three of these staff noted that they collected this information for their own purposes and that the data was never compiled in a digital record, presented to their team or management, and that it served the purpose of generating new topics for presentations. In order words, many programs use various methods to collect data about their programs but do not carry out other steps of program evaluation, such as systematically analyzing data, summarizing and reporting findings. Another coordinator mentioned that they developed a client follow-up process to check-in on client progress but that other teams did not know of the practice and the information was never used for any other purpose.

The Language and Vocational Assessment program participated in a long-term outcome measurement project with their provincial funder, which was led by an external evaluator. The project led to increased use and mainstreaming of evaluation into the program's day-to-day practices. The program developed a new tool for collecting intermediate outcomes that was incorporated into frontline staff's day-to-day responsibilities. The information gathered was then used to secure resources during a period of funding cuts across the province. Furthermore, evaluation became a standing agenda item during the program's team meetings to create buy-in for the new practice with frontline staff, communicate the importance of the new practice and support staff exchange and learning about evaluation. The discussion participants from this program provided an example of evaluation being mainstreamed into day-to-day activities, such as team meetings, and use of the data for accountability and funding, capacity building and organizational learning.

4.3.3 Barriers to Evaluation

The barriers to evaluation include the nature of working with newcomer participants, including language, culture and power differentials between staff and participants; funder leadership and guidance for evaluation and the subsequent learning climate; resources for evaluation, such as time and evaluation staff; and, challenges to mainstreaming evaluation as a result of leadership and resource constraints.

Program Participants

Participants across the three groups perceived that the surveys and focus group methods currently carried out were inaccurate, lacked integrity, or were not useful due to language and literacy levels of clients, cultural barriers, and power dynamics between clients and staff. Participants made the following comments related to language barriers:

- "We have our programs directly with immigrants, newcomers, refugees, and the evaluation depends on the literacy level of the person. We have a standard evaluation that I don't think works for every single person… I think the evaluation only works with

people who have a specific level of literacy and they are able to understand our questions. Language is a really big issue in our evaluation. When language is a big issue they are not able to answer our questions and they have very short answers. They just give a very very short answer and they have the tendency to copy one another, pre-Covid times. “What did you say? You said it was good? It works? Yes it works!” It becomes something that we cannot use.”
- “To get around the literacy portion, we do it with more advanced clients, CLB 3 and higher, but we don’t ask true beginners to take part in the survey because like [the other participant] said, it’s just not effective.”

Participants expressed that, although interpretation and translation is often provided for clients, the agency relies on internal staff support workers to interpret focus groups or translate surveys. Therefore, two participants noted that it is difficult to record meaningful feedback or measure program outcomes because clients’ responses are mediated or filtered through staff.

Through the small group discussion process, participants reflected on the cultural aspects of evaluation and how the process of completing surveys or participating in focus groups may be strange or uncomfortable for clients. Participants commented about how culture impacts evaluation efforts:

- “But in other cultures the response is extreme “No I hated it, I didn’t learn anything” or “Yes, this was the most fantastic thing in the world, you were perfect, everything is perfect” so it’s hard to make the feedback beneficial.”
- “And with the newcomers it becomes an awkward scenario, after building trust, putting a piece of paper in front of them and having to explain it. Frontline workers don’t have time, find it awkward and all of the above. Clients are typically thankful for receiving services and participating in programs. They just want to be thankful and don’t want to say something critical. In my program, we’d have to keep saying no really it’s useful for us to know. But they’ll never give you the truth, they’re too thankful, and have a different sense of respect when it comes to engaging and participating in this kind of thing. And [evaluation] is not a part of many newcomers’ cultures in my experience.”

Finally, participants discussed power differentials between staff and clients as a barrier to evaluation:

- “When clients attend [my program], they’ve only been in the country for a month, so everything is brand new including us. We have this intimidation, there is a hierarchy feel and they don’t know exactly who we are and our relationship to the government and what they should do. So they’re more keen on copying each other and not being honest in their evaluation.”
- “And then you have the cultural aspect “you want me to evaluate?” because they don’t feel they’re in a position of power to be honest.”
- “In our programs, some of our clients are sent by the court or judge to attend. So they don’t want to say anything. We try to tell them it’s a safe space, we’re not the government, but they’re involved because of child abuse and issues with their children.

So they will never say anything that is critical. Because they think we have a power we don't have and they are afraid of the system."

- "There is certainly a hierarchy because the client would perceive that we would take away services or provide lesser services if they didn't answer correctly. But it's hard to get over that, especially for clients with lower education or cultures with strict hierarchy."

Participants questioned whether it was possible to reduce power differentials sufficiently to accurately measure the outcome of their work and to receive feedback for how to improve services. Taken together, language, culture and issues of power between clients and staff present a significant challenge to gathering both summative and formative program data from clients and subsequently evaluating services.

Leadership and Learning Climate

An additional barrier that was identified both explicitly and implicitly in the small group discussions is that evaluation is understood by staff at all levels as a mechanism of accountability towards funders. However, staff showed less understanding of the other benefits of evaluation, namely as a tool for learning, social innovation and engagement. Participants at the Coordinator and Team Leader level explored the organizational perception of evaluation as accountability and wondered how evaluation could be used to learn and grow within the organization, while the conversation among Managers mostly framed evaluation more exclusively for accountability. One participant noted that the message passed down from management is that the questions of highest importance in surveys and focus groups are those that relate explicitly to federal outcomes:

- "They were 5 – 10 questions and we were told that only 2 of the questions were for the funder and the rest of them don't really matter. And when I tried to pursue changing the survey, there was a lot of rigidity like "no we can't change it! We've done it that way forever" but we're only using two of the questions."

The Coordinators particularly discussed the other benefits to evaluation and that the organization tends to focus on accountability, but that they have the motivation to pursue evaluation that supports improvement and program development, both of which would benefit clients. One Team Leader spoke about how evaluation could better meet the needs and interests of both staff and funders and that the challenge is "finding the happy medium."

All three participant groups discussed how the relationship with funders impacts evaluation, long term planning and the culture of evaluation within the service. Many participants also discussed the challenging funding climate of their work and how that impacts program evaluation. One Team Leader noted the difficulty in implementing evaluation and evidence-based decision making and strategic planning in a volatile funding climate, particularly for programs with short-term contracts. At the Management level, there was discussion about the lack of clarity from the federal funder about evaluation expectations and scope, articulated in the following comment:

- "So maybe there should be a bigger conversation about how each particular service, or how an outcome can be revisited or revised, based on what my service is doing to make a difference in someone's life. The outcomes are very broad and big and any service in the end can be contributed to the ultimate client success or lack thereof. There should be more conversations about more realistic outcomes that my service can report on without [IRCC] asking "how do I know it's your service that made an impact?" because that's what it boils down to. If we report on what they ask us to report on, they then respond "well how do we know it's your service that caused that?" And it's probably a combination of multiple services, so maybe we need to make it more specific."

This participant touches on the fact that the IRCC defines the outcomes for the programs they fund, yet the outcomes they provide are broad and can be attributed to a multitude of factors in the client's life. Therefore, the participant questioned what the role and responsibility is of the federal funder in adjusting outcomes so that they are more feasible and realistic for individual agencies to measure.

Resources

The two resources that came up across the three participant groups were lack of time and a lack of staff or internal department specifically dedicated to evaluation. Each of the three participant groups discussed the real or perceived lack of time to carry out evaluation due to the urgent nature of frontline work, a culture of busyness and that evaluation is not prioritized as a result. Participant comments on the topic of time include:

- "The logic model and evaluation was put on top of a system already running, so it seemed like extra work. Same with Settlement Counsellors. It seems like extra work. In the nonprofit sector we're always in over our head. And staff just don't have the time. Getting the survey we need for our reporting is super hard."
- "We have a culture of busy-ness and evaluation is used to prove that we are busy. It's not a tool that is incorporated in the culture because we instinctively want to prove that we are busier than anyone, and I think we have our blinders on. We want to feel busy and want to focus on where we're comfortable."

Participants discussed time as a *real* barrier in that many programs work with clients in crisis on a daily basis who have urgent needs that must be met. As a result, current evaluation tools become an additional administrative burden on frontline staff, such as completing a paper survey with clients or inviting clients to a focus group. However, participants also discussed time as a *perceived* barrier because staff's desire to appear busy, possibly due to fears around job security in an unstable funding climate. Furthermore, participants cited that they are carrying out tasks in their day-to-day work that they find meaningful, such as checking-in with clients and completing informal follow-ups, which could be shifted to being dual-purpose for evaluation. However, because these tasks were not specifically requested by management they were not perceived as evaluation. Furthermore, these comments reveal that staff have the time to complete the tasks they find meaningful and add value to their work, but perceive evaluation that is disconnected from their work and what they find meaningful as a burden on their time.

Team Leaders and Program Managers discussed the lack of an evaluation department, which impacts evaluation follow-through, knowledge sharing, mainstreaming, and the ability to make evaluation more meaningful in the organization. Their comments include:

- "But the real evaluation of everything and writing an evaluation report never happened. There was no time. We look at it but not in the bigger picture. And I think it would be helpful to have someone take on evaluation in the agency to do that."
- "I think the way to solve the frontline competence piece is to really offer support and assistance with a designated staff from within to help with that. So that our excellent staff doing really important work don't get overwhelmed but it contributes to their success and continued funding."
- "I think there should be a team, even two people. It's a complex and time consuming topic, interesting, useful for every piece of our work. Even to show to the volunteers the impact of their work, thanks to an evaluation or as a survey or a focus group. For simple things, to get what we want from the community: volunteers, funding, grants, to improve our services and what needs to change and improve. It's not a job you can do half day a week. There are a lot of things to analyze and we've grown so much in the last few years. It's a job for two people, if we want to do something good."

Furthermore, participant comments regarding time and an internal evaluator show how these two issues of resources intersect because more meaningful evaluation requires significant time, which could be filled by an additional position or research team.

4.3.4 Facilitators to Evaluation

There are a number of strengths within the organization related to evaluation, as evidenced in the small group discussions, including: a breadth of evaluation practices; knowledge, experience and training in evaluation; and, staff who are pursuing evaluative activities in their own time for the purpose of learning and improvement.

Breadth of Methods

As highlighted in Section 4.3.1 Doing Evaluation, ISS uses a breadth of data collection methods, including individual program surveys, focus groups, client follow-ups, and interviews that encompass aspects of both summative and formative evaluation. As evidenced by the discussion participants, the organization has experience gathering data for accountability and improvement purposes. Some participants also were able to distinguish how specific questions related to measuring outcomes as defined by a funder and those that were for internal improvement purposes. Furthermore, some participants spoke of supporting external evaluators to evaluate their programs, which brought additional evaluative practices and knowledge into the agency. One participant who had worked with an external evaluator discussed their interest in pursuing more creative evaluative practices, such as photovoice, as a means to engage low language and literacy clients.

Evaluation Training and Experience

All three groups discussed their knowledge and training in evaluation, both formal and informal, and how their programs are currently evaluated. The Coordinators all had experience distributing surveys to their clients, inviting clients to participate in focus groups and interviews for evaluation purposes and one had attended a formal evaluation training session through the organization. All participants at the Team Leader and Program Management level had attended training on evaluation and outcome measurement, including one who had taken a graduate course in project implementation and evaluation and the majority had attended short seminars on outcome measurement for social service providers. The recency of training for leadership staff ranged from some participants being currently enrolled in an outcome measurement course remotely, a couple who had taken training in the last two years, and others who had taken an outcome measurement training more than ten years ago. Two Team Leaders had cited that they either received training or participated in a long-term evaluation project as required by a provincial funder.

Participants Evaluating Programs Informally

Lastly, as was discussed in the previous section, participants described their own means of obtaining information about their clients for program improvement and learning, but they did not necessarily equate this practice with evaluation. Despite participants not considering these practices formally as evaluation, they still signal that participants are interested in gathering meaningful data that will enrich their work and ultimately benefit clients. The comments include the following:

- "I personally as a facilitator, in developing the sessions, I made slides every day about resources in the community. It resonated with me as a facilitator. It didn't come from an evaluation or following trends over a period of time. It came because I saw it and I saw it as a need and I decided personally to do it. There is no database supporting the decision."
- "Things are often self-driven for surveys. Even in [program], we would do six month follow-ups... And no one else knows that this work is happening outside our team, which speaks to the siloed nature of our work. We have such great skills across the board. It would be great to have someone overseeing it and also having a mechanism to share the information with other staff."

Therefore, the challenge is to find a means of balancing accountability demands and the interests and motivations of staff and information that will enrich their work with clients.

4.4 Integration of the Methods

The ECAI survey provided an opportunity to measure evaluation capacity at ISS in the context of individual factors, organizational factors and evaluation capacity outcomes. The small group discussions allowed for further exploration into the overall scores from the ECAI by sub factor and contextualizing the scores from the participants' perspectives and understanding of evaluation capacity and practices at ISS. Table 6 integrates findings from the ECAI with corresponding findings from the small group discussions in an effort to integrate the results from both methods (Creswell and Plano Clark, 2017, p. 239).

Table 6. Joint Display of ECAI and Small Group Discussion Findings

ECAI Scores	Small Group Discussion Major Findings
Individual Factors	
Awareness (3.31)	● Strong agreement among participants of the importance of evaluation, particularly for accountability and improvement purposes.
Motivation (3.59)	● Interest in learning more about evaluation and integrating evaluation and outcome measurement practices across ISS programs. ● Interest in learning new ways that evaluation could benefit ISS and ultimately service delivery and client outcomes. ● Interest in learning more about evaluation for learning and development purposes and approaches that are language and culturally accessible to clients, such as creative approaches.
Competency (3.06)	● Most participants had a mix of formal and informal experience evaluating their programs and taking evaluation related training, although training ranged in recency from current enrollment to ten years ago. ● Mention of inability, either due to time or competency, to systematically analyze data that is already being collected.
Organizational Factors	
Leadership (3.05)	● General confidence in the leadership provided by upper management. ● Perception that communication from management related to evaluation is heavily focused on accountability and less for internal learning. ● Coordinators and Team Leaders experience evaluation activities as disjointed due to big pushes in evaluation data at reporting periods and little other communicated uses for information outside of reporting cycles. ● Sense of separation and siloes between programs.
Learning Climate (2.91)	● Respect for others and ability to share knowledge and experience with colleagues. ● Expressed culture of compliance with funder and accountability expectations - discussions at all three levels focused on compliance with funder expectations, which is at odds with a climate of learning and experimentation.

Resources (2.98)	● Real and/or perceived lack of time to complete evaluation, analyze data and relay findings to staff. ● Lack of internal evaluation staff, department or committee, which could create cohesion among evaluation efforts across programs. ● Lack of resources, leadership and guidance from the major funder related to evaluation and outcome measurement expectations and practices.
Evaluation Capacity Outcomes	
Mainstreaming (3.09)	● One example of evaluation mainstreaming into team meetings and the daily work practices of a program. ● Evaluation expressed as a requirement at reporting periods but not integrated into day-to-day practices for the majority of participants.
Use of Findings (2.88)	● Current program data collected is used predominantly for internal and external accountability in written reports and program improvement. ● Minimal mention of uses of evaluation for learning, staff engagement, or longer term planning.

The ECAI and small group discussion findings show complementary results. The organization has a solid foundation of data collection methods across programs, could benefit from strengthening and updating evaluation competencies, and there is an interest in such opportunities. However, the primary barrier to evaluation capacity is that evaluation is practiced as a compliance with accountability requirements, rather than as a learning, engagement and strategic function, which was evident in the survey but especially highlighted in the small group discussions. The evaluative activities at ISS are heavily impacted by funder relationships, which limits its use within the organization. Although a breadth of data is collected across programs, the data has a short lifespan when it is used singularly for reporting and accountability functions. The short life span of data and its limited use negatively impacts frontline staff's commitment to evaluation, such as lack of time in their day to dedicate to evaluation because of its limited applicability to their day-to-day work. Essentially, evaluation capacity outcomes, such as evaluation use and mainstreaming, are limited because organizational resources for evaluation are lacking. Resources for evaluation include both financial resources and leadership and guidance for evaluation from funders.

The organization could benefit from experimenting with new approaches to evaluation, such as collaborative approaches to evaluation, that involve frontline staff in the evaluation cycle, including in the following processes: defining internal outcomes that are not dictated by the funder; determining how and from whom the data will be collected; analyzing and reporting results to colleagues; and, collaboratively determining how the results will be used. Collaborative approaches to evaluation projects could increase frontline competencies in evaluation, support internal learning about evaluation, and increase the likelihood the evaluation is meaningful and will be used. Furthermore, participating in such an initiative could also foster relationships between staff of different programs and support greater understanding of how everyone's duties fit together with the broader mission of the service, which was also identified as an area of improvement in the ECAI.

5.0 Discussion and Analysis

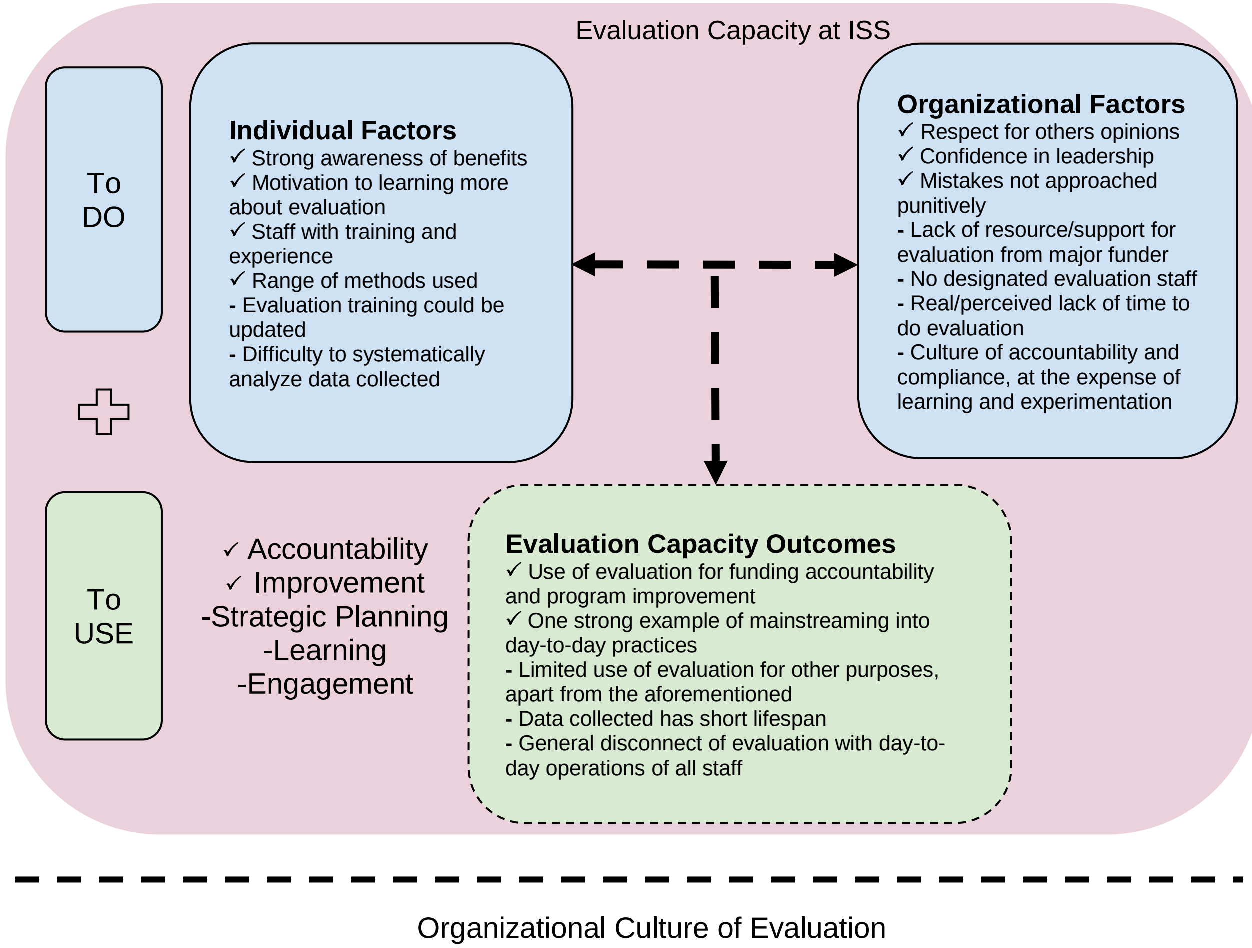
The discussion and analysis in this section will return to the research objectives and questions presented in section 1: understanding the evaluation capacity at ISS, better understanding frontline staff, Team Leaders and Program Managers' experience with evaluation and exploring how the organization can develop a culture of evaluation. The discussion will be embedded within the conceptual framework presented, which was developed through a review of the literature on accountability, evaluation capacity and nonprofit evaluation practices.

5.1 Evaluation Capacity at ISS

The ECAI used for the quantitative phase of the research does not assign a specific degree of evaluation capacity based on the score achieved. Rather, evaluation capacity is explored through the factors of individual, organizational and evaluation capacity outcomes of the agency. These factors inform the organization's ability to do and to use evaluation, which ultimately contribute to the organization's culture of evaluation. Bourgeois and Cousins (2008) developed another tool for government agencies that places organizations at a stage of evaluation capacity based on an empirical assessment. In their tool, organizations can have exemplary, intermediate, developing or low capacity (p. 139). According to the findings of this project, the ISS has developing capacity for evaluation. The evaluations completed are at the individual program level, are mandated externally and primarily pursued and used for external accountability, there is not an evaluation team, and there are varying levels of technical evaluation capacity of individuals. Evaluation activities are pursued for individual programs and do not connect to the wider mission of the service or agency. However, evaluative activities are completed on a regular basis, there is a breadth of evaluative information collected and methods used, and information is used for accountability and improvement purposes.

Figure 10 integrates the findings from the previous section into the conceptual framework provided in Section 1.4. The points in the figure below with a '✓ indicate a facilitator to evaluation capacity and those with a '-' indicate a barrier to evaluation capacity and potential area of future focus for evaluation capacity building interventions. There are facilitators and barriers presented in each of the evaluation capacity factors, with the most significant barriers being presented in the organizational factors. The issues present in the organizational factors, such as lack of evaluation leadership, resources, and guidance from a major funder, limit the organization's ability to prioritize and use evaluation for learning, engagement and growth, and subsequently develop a culture of evaluation.

Figure 10. Evaluation Capacity at ISS within Conceptual Framework



5.1.1 Individual Factors

Staff are aware of the benefits of evaluation and motivated to gather information that will benefit their clients. In some cases staff are carrying out evaluative activities, such as client follow-ups, informally because they want to know how their clients are doing and how to improve. Evaluative experience and competency is not necessarily a deficit at ISS, but there is a need to update skills and knowledge of evaluation, including understanding and utilizing additional approaches to evaluation. Evaluators should be flexible in various data collection methods and evaluation approaches in order to fit the program context as the appropriateness of methods and approaches is dependent on the context, nature and goals of a program or intervention (McDavid, Huse and Hawthorn, 2017). Furthermore, although one-off training in evaluation is attractive to management both in terms of time and financial resources, they are not necessarily effective in building long term evaluation capacity (Despard, 2016). Rather, agencies should undertake a variety of approaches to build individual capacity, including communities of practice, mentorship (Preskill and Boyle, 2008), and community collaboration, such as with fellow service providers and university-community partnerships (Bakken, Núñez, and Couture, 2014). According to Despard (2016) the most effective evaluation capacity building intervention is one-on-one technical assistance over several months, such as internal staff working alongside an experienced evaluator to develop and deliver an evaluation plan. This type of evaluation capacity building intervention could greatly impact the individual competency for evaluation of a small number of staff.

5.1.2 Organizational Factors

Leadership

The topic of evaluation for accountability was evident in both the ECAI and the small group discussions. In the ECAI, participants rated evaluation use for funders as the highest rated use statement, which was then reflected in the small group discussions across all three participant groups. Evaluation is primarily understood and communicated about in the organization as a tool of accountability to funders, although there is interest in pursuing evaluation for other purposes. The manner in which evaluation is described and communicated about impacts its quality and use. As reflected in the literature, focusing heavily on accountability and performance measures limits evaluative learning, inquiry, exploration and ultimately use (McDavid, Huse and Hawthorn, 2017; Cousins, 2019). In addition, focusing too heavily on accountability and performance measures is bad for morale and ultimately decreases staff motivation to comply with such measures (McDavid, Huse and Hawthorn, 2017). This point could be particularly pertinent for ISS, and is worth exploring more, as a workforce who largely are in humanitarian positions because of personal convictions, beliefs and desire to make a positive impact. Unfortunately, referencing funder accountability to incite compliance does not necessarily lead to performance measurement and evaluation buy-in from frontline staff, so other communication tools and vernacular should be employed that are meaningful to staff and speak to their motivation for working in the organization (Perez Jolles, Collins-Camargo, McBeath, Bunger and Chuang, 2017). The organization should develop ways of communicating evaluation objectives with staff that better balance accountability with learning, improvement and positive client outcomes, which is likely to connect more deeply with frontline workers (Liket, Rey-Garcia and Maas, 2014).

Compliance and accountability are not inherently negative qualities of an organization. However, compliance and heavy focus on upward accountability to funders can have negative impacts on organizations when paired with funding dependency, because it puts nonprofits in a position to orient activities to maximize government contracting rather than focusing on the needs of clients and gaps in services experienced by communities. Ultimately, compliance and accountability are often placed above the organization’s desire to learn, adapt and socially innovate. From a leadership perspective, it might be worth considering if ISS is already sufficiently accountable to funders and at what point evaluative activities can shift to better balance the interests of multiple stakeholders, including funders, staff and clients (Chouinard, 2013). As noted by Ebrahim (2016), stringent and ever-increasing accountability requirements placed on nonprofits, particularly those with funder relationships marked by financial dependence, can have negative impacts on an organization’s ability to achieve their social mission and build capacity. Therefore, the goal should not be to ignore contracting demands for accountability, but rather to equally prioritize accountability with organizational learning.

Learning Climate

There are currently little internal platforms for evaluative inquiry and information at ISS, which means that apart from reports to funders, there are few avenues to convey the results of evaluation activities and practice. These topics can be incorporated into staff meetings and regular presentations to increase staff knowledge of the purpose of these activities, and subsequently increase staff capacity for evidence-based thinking and practice. Another option is to incorporate evaluation and research conducted internally into regular staff development opportunities. Without such avenues, evaluative practice will continue to be disjointed and disconnected from frontline staff's day-to-day operations and service delivery practices. Nonprofits tend to lose sight of evaluation and long-term planning as they meet the urgent demands of clients and funders (Gill, 2010), but the continual focus on urgent daily tasks challenges organizational capacity and can negatively impact the ability to enact long-term systems change. Evaluative practices and thinking provide opportunities for organizations to collect, analyze, and reflect on information related to their day-to-day work while also considering their next steps and how to make longer-term impacts on clients' lives and communities.

Resources

The issue of leadership and support for evaluation provided by a funder at ISS was prevalent in both of the ECAI statements related to resources for evaluation and in the small group discussions. In a report published by Broughton and Shields (2020), the issue of instrumental upward accountability, or contract management, is present across the immigrant settlement sector in Canada and is not isolated to the case study of ISS. According to their report, the federal government's emphasis on instrumental accountability and contract management creates a culture of compliance and evaluative 'box checking' in funded agencies, which ultimately challenges service provider capacity, rather than building and enhancing it. The onerous requirements placed on agencies are found to detract from their missions of supporting the settlement of newcomers and, in the end, negatively impacts their ability to focus on meeting clients' needs and supporting successful client outcomes. Broughton and Shields's findings are reflective of the present case study of ISS, as an IRCC-funded settlement service provider, and demonstrates how a funder's approach to evaluation and performance measures can trickle down to agency leadership and frontline staff's understanding of evaluation and its use.

One resource that is not included in the ECAI but was mentioned in the small group discussions is frontline staff. The ECAI includes staff knowledge and competency for evaluation in the individual factors, but in the context of ISS staff should be considered a resource for linguistic and cultural knowledge that can inform the improvement of evaluative practices. Currently, the funder is communicated as the primary stakeholder and consumer of evaluative information, which is demonstrated through interim and annual reports. However, frontline staff should be providing consultation on current practices to ensure they are accessible and applicable to the context of their clients and they should also be aware of how the information is used. The organization should be responsible for reporting back the results of an evaluation to frontline staff and closing evaluation feedback loops, as is done with funders, because they have a high interest and stake in the success of their programs (Gill, 2010). Not only should frontline staff be included in developing methods and the collection of data, but they should be keenly aware of how the data is used and what the results of an evaluation of their program are. Increased involvement of the frontline staff of ISS in the evaluation process would increase relevance of data gathered, ownership and follow-through of the evaluation practices within the agency, and opportunity for internal learning about program and client context (Cousins, 2019).

Another aspect of resources that was significant in both the ECAI and small group discussions was resources to facilitate client participation in evaluation, such as interpretation and translation. Participants in the ECAI indicated that resources are provided to facilitate client participation in evaluation, but the small group discussions lent additional context into the realities of interpreted and translated methods. In the small group discussions, participants stated that regardless of the interpretation being provided, they still did not feel like the information gathered was meaningful because of the cultural differences and power differentials between staff. The use of the mixed methods in this research provided an opportunity to delve into those themes, and draw attention to the need for the organization to explore more culturally responsive approaches to evaluation (Chouinard and Cram, 2020). In the context of ISS, a culturally responsive approach could include working with frontline staff who have lived experiences as newcomers, as mentioned above, to collaboratively develop evaluation tools that are appropriate to the language and cultural contexts of clients. Furthermore, frontline staff may be able to inform how the organization might address power differentials in the evaluation process, although it may not be possible to eliminate them (Cornwall and Jewkes, 1995).

Lastly, as noted significantly in the small group discussions, ISS lacks a designated staff or department specifically for evaluation. A designated staff or team could support a number of the barriers to evaluation capacity that have been described prior, such as carrying out evaluation capacity building training and initiatives, supporting staff to strengthen evaluative practices, following through with data collection and analysis, reporting findings back to staff, and following-up on evaluation use. At ISS, the largest information bottleneck takes place after data has been collected. Large amounts of data are collected that staff then do not have the time of competency to systematically analyze, but this issue could be solved with a designated evaluation staff or department. Furthermore, culturally responsive and collaborative approaches (Chouinard and Cram, 2020; Cousins, 2019) aim to build the collective capacity of non-evaluators and increase evaluation use within organizations. However, these approaches still require a designated person to act as facilitator, guide or catalyst of the evaluation initiative. Additionally, a designated internal evaluator also ensures that data quality and evaluation integrity are being maintained despite using a collaborative approach with staff who are not trained in research or evaluation (Cousins, 2019; Fetterman, 2001).

5.1.3 Evaluation Capacity Outcomes

Currently, the evaluative data ISS collects has a short lifespan, which ultimately means it is an expensive practice with little utility. As indicated in the small group discussions, qualitative data, such as collected in client focus groups, is used by pulling quotes to include in government reports. However, focus groups and other qualitative methods are expensive and labour intensive when considering the resources invested in collecting such information, mostly in the form of paid staff time: frontline staff hours to call clients to invite them to participate; focus group facilitator hours; community support worker hours who provide interpretation; and, the hours to transcribe the focus groups. The use of quotes in a report, compared to systematically analyzing data according to a methodology and reporting findings, means that a large amount of paid staff time is being invested in a practice that has extremely limited use. This is not meant to imply that the practice should be ended and replaced with a more cost effective practice, but that if the paid hours are already being invested then they should be maximized through multiple uses, such as contributing to ongoing learning and development, staff engagement, and strategic planning.

5.2 Culture of Evaluation

As is evident in the conceptual framework, a culture of evaluation is the ultimate goal of building evaluation capacity. A culture of evaluation, as defined by Mayne (2008), is one that engages all staff in self-examination and reflection, uses evidence-based learning and practices, and promotes experimentation and innovation (p. 5). As discussed in the introduction, this classification of organizational culture is beneficial to nonprofits who operate in increasingly competitive fiscal environments and who deal with increasingly complex social issues. ISS experiences challenges to evaluation capacity that are barriers to achieving a culture of evaluation, which are highlighted in the previous sections. ISS could benefit from dedicating time and spaces to reflect on the work they do as an organization with all staff. Holding space for reflective and strategic practices should not be limited to leadership staff, but include all staff in regular intervals. Furthermore, without a research or evaluation-focused staff or department, the organization continues to collect data that they do not have the capacity to systematically analyze, summarize, and report back to staff and other stakeholders, which could result in greater learning opportunities and capacity building. Lastly, while the organization has to meet the accountability requirements placed on them by funders, they should also explore evaluative activities that will enrich the organization internally for their own purposes, such as community-based research and collaborative approaches to evaluation with frontline staff and clients, which would lead to learning about gaps in services and developing new and innovative programs and practices. Staff have a strong awareness of the benefits of evaluation, are motivated to learn more about it, and already have a strong foundation in evaluative activities within the agency, which will be an asset with evaluation capacity building initiatives moving forward.

5.3 Additional Implications

ISS experiences barriers to evaluation that are typical of nonprofits, as found in the literature review, such as lack of: financial resources and leadership from funders; time to complete evaluation; and, skilled evaluation staff or a dedicated evaluation department. Furthermore, ISS experiences additional barriers as an immigrant serving agency due to the nature of working with newcomer participants, such as language, culture and power imbalances between staff and clients. Additional research is needed on how newcomer serving agencies across Canada evaluate their services, particularly in light of Broughton and Shield's (2020) report which indicates the environment of accountability and contract compliance present in IRCC-funded agencies. This research could take the form of a provincial scan of IRCC-funded agencies in Alberta or nationally to determine their shared challenges to building evaluation capacity across the sector and how agencies are evaluating their services despite shared barriers. Unfortunately, although IRCC has funded national projects to build evaluation capacity of the sector, these projects have largely taken the form of one-off evaluation training, webinars or modules, which only address individual competency for evaluation and not organizational factors or evaluation capacity outcomes. As a major funder, it may be more effective to fund internal evaluators either within agencies or in provincial umbrella organizations and then leverage learnings across the country through provincial or national communities of practice.

6.0 Recommendations

6.1 Introduction

The following section outlines three recommendations for ISS to build their evaluation capacity, which can be adopted in whole or in part by the organization: investing in evaluation, such as having an internal evaluator on staff; balance intended uses of evaluation; and, explore additional approaches to evaluation, such as collaborative and culturally responsive approaches in future evaluations.

6.2 Recommendations

6.2.1 Invest in Evaluation

First, ISS should consider investing in an internal evaluator to harmonize research and evaluation practices within the service. An internal evaluator could serve multiple purposes within the service, in additional to actually carrying out evaluations, such as:

- Build internal capacity by planning and delivering evaluation training with staff;
- Support program and project development, including strengthening outcome measurement frameworks in funding proposals;
- Address data bottlenecks and ensure that evaluation data is compiled, analyzed, and reported back to stakeholders, including staff;
- Build or maintain agency relationships with higher education institutions and discovering and vetting research projects with academics and students; and,
- Spearhead evaluative projects that actively involve staff, such as collaborative and culturally responsive approaches and creative methods, such as client photovoice projects and journey mapping.

6.2.2 Balance Intended Uses of Evaluation

ISS should attempt to better balance evaluation for accountability with other organizational functions to maximize its utility, including uses for accountability, training, learning, staff engagement, strategic planning and program development. Such a shift could include adjusting current evaluative tools to balance questions relating to contract outcomes, client experiences, gaps in community services, and/or addressing agency mission and vision. Alternatively, ISS can develop an evaluation calendar that includes funder accountability evaluations and internal learning and development evaluations. Additionally, time and resources can be dedicated to evaluative activities that are more creative and exploratory, such as photovoice or journey mapping with clients. These activities can be pursued by an internal staff or by leveraging relationships and assets in the community, such as graduate student research projects (Bakken, Núñez, and Couture, 2014).

6.2.3 Explore Collaborative and Culturally Responsive Approaches

A final recommendation is to increase the involvement of frontline staff in evaluative activities to increase relevance of data gathered, program and data ownership, follow-through of the evaluation practices within the agency, and opportunity for learning about program and client context (Cousins, 2019). If staff have a say in what data is meaningful to them, their work, and their clients, it will have a positive impact on the overall utility of evaluation and increase platforms for evaluative inquiry and information. Evaluation and internal research projects can contribute to staff development opportunities, team and all-staff meetings, and agency presentations at conferences, among others. One example of a collaborative approach could be to have staff support the evaluation of an internal committee or project throughout the evaluative cycle, which would result in learning about the internal initiative and the evaluative process. Furthermore, ISS should explore approaches to evaluation and methods that are more responsive to the cultural context of clients, such as a culturally responsive approach (Chouinard and Cram, 2020), given the barriers to evaluation that are present in the service and sector generally.

7.0 Conclusion

The research undertaken aimed to understand the evaluation capacity of the Immigration and Settlement Service of Catholic Social Services, including staff experience and understanding of evaluation, and explore their barriers and facilitators to build evaluation capacity. The conceptual framework, developed from the literature review, explored evaluation capacity through the lens of individual, organizational, and evaluation capacity outcome factors, which describe an organization's ability both to do and to use evaluation. Through the use of the Evaluation Capacity Assessment Instrument (ECAI) and small group discussions, the researcher was able to identify specific strengths and limitations of the agency's current evaluative practices. In terms of strengths, ISS staff are aware of the benefits of evaluation, are motivated to learn more about evaluation, and there is already a breadth of evaluative practices taking place within the organization. The organization also faces a number of barriers to evaluation, such as the nature of working with newcomer participants, including differences in language, culture and power, lack of time and internal evaluation staff, and lack of support and guidance for evaluation for their major funder. These issues are a barrier to the organization developing a culture of evaluation. The organization would benefit from investing financial resources in an internal staff evaluator, better balancing the uses of evaluation, such as for accountability, learning and staff engagement, and exploring additional approaches to evaluation to promote use, such as collaborative and culturally responsive approaches.

References

Acree, J., & Chouinard, J. A. (2019). Exploring use and influence in culturally responsive approaches to evaluation: A review of the empirical literature. *American Journal of Evaluation,* , 109821401987950. doi:10.1177/1098214019879505

Almog-Bar, M. (2018; 2017). Civil society and nonprofits in the age of new public governance: Current trends and their implications for theory and practice. *Nonprofit Policy Forum, 8*(4), 343-349. doi:10.1515/npf-2018-0004

Bach-Mortensen, A. M., & Montgomery, P. (2018). What are the barriers and facilitators for third sector organisations (non-profits) to evaluate their services? A systematic review. *Systematic Reviews, 7*(1), 13-15. doi:10.1186/s13643-018-0681-1

Bakken, L. L., Núñez, J., & Couture, C. (2014). A course model for building evaluation capacity through a University–Community partnership. *American Journal of Evaluation, 35*(4), 579-593. doi:10.1177/1098214014523671

Behrens, T. R. (2011). In Agard K. A. (Ed.), *Role of nonprofit leaders in data and needs analysis and assessment*

Botcheva, L., White, C.R., & Huffman, L.C. (2002). Learning culture and outcomes measurement practices in community agencies. *American Journal of Evaluation, 23*(4), 421-434. doi:10.1016/S1098-2140(02)00229-1

Bourgeois, I., & Cousins, J. B. (2008). Informing evaluation capacity building through profiling organizational capacity for evaluation: An empirical examination of four Canadian federal government organizations. *The Canadian Journal of Program Evaluation, 23*(3), 127

Bourgeois, I., & Whynot, J. (2018). Strategic evaluation utilization in the Canadian federal government. *Canadian Journal of Program Evaluation, 32*(3) doi:10.3138/cjpe.43179

Bourgeois, I., Whynot, J., & Thériault, É. (2015). Application of an organizational evaluation capacity self-assessment instrument to different organizations: Similarities and lessons learned. *Evaluation and Program Planning, 50*, 47-55. doi:10.1016/j.evalprogplan.2015.01.004

Brinkerhoff, J.M. (2002). Government-Nonprofit Partnerships: A Defining Framework. Public Administration and Development 22, 19–30 (2002)

Broughton, S. and Shields, J. (2020 April). Resilience and the Immigrant Settlement Sector: A Consideration of the Place of Accountability and Performance Measurement. York University. Accessed on January 10, 2021 from https://bmrc-irmu.info.yorku.ca/files/2020/04/April-2020-Shields-Full-Report-FINAL.pdf?x82641

Canadian Evaluation Society (2015 October). What is evaluation? Accessed on January 21, 2021 from https://evaluationcanada.ca/sites/default/files/ces_def_of_evaluation_201510.pdf

Carman, J. G., & Fredericks, K. A. (2010). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation, 31*(1), 84-104. doi:10.1177/1098214009352361

Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation, 28*(1), 60-75. doi:10.1177/1098214006296245

Chaytor, K., MacDonald, G., & Melvin, D. (2002). Preparing nonprofits for new accountability demands. *The Canadian Journal of Program Evaluation/La Revue Canadienne d'Evaluation De Programme, 17*, 93-112.

Chelimsky, E. (1997). The coming transformations in evaluation. *Evaluation for the 21st Century: A Handbook*. Thousand Oaks: SAGE Publications, Inc. doi:10.4135/9781483348896.n1

Cheng, D. W. H. (2003). *Evaluating the evaluation: Assessing the use of program evaluation for capacity building in a nonprofit organisation* (M.E.Des.). Available from ProQuest Dissertations & Theses A&I. (305347859). Retrieved from http://search.proquest.com.ezproxy.library.uvic.ca/docview/305347859?accountid=14846

Chouinard, J. A., & Milley, P. (2015). From new public management to new political governance: Implications for evaluation. *The Canadian Journal of Program Evaluation, 30*(1), 1-22.

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation, 34*(2), 237-253. doi:10.1177/1098214013478142

Chouinard, J. A. and Cram, F. (2020). *Culturally Responsive Approaches to Evaluation.* SAGE Publications.

Compton, D. W., Baizerman, M., & Stockdill, S. H. (2002). *The art, craft, and science of evaluation capacity building*. San Francisco: Jossey-Bass.

Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation, 18*(1), 61-77. doi:10.1177/1356389011430371

Cornwall, A., & Jewkes, R. (1995). What is participatory research? Social Science & Medicine (1982), 41(12), 1667-1676. doi:10.1016/0277-9536(95)00127-s

Cornwall, A. (2008). Unpacking 'participation': Models, meanings and practices. Community Development Journal, 43(3), 269-283. doi:10.1093/cdj/bsn010

Cousins, B. (2019). *Collaborative Approaches to Evaluation: Principals in Use*. SAGE Publications.

Cousins, B., and Whitmore, E. (1998). New Directions for Evaluation. Jossey-Bass Publishers: no. 80. Winter 1998.

Cousins, J. B., & Bourgeois, I. (2014). *Organizational capacity to do and use evaluation : New directions for evaluation, number 141*. Somerset: John Wiley & Sons, Incorporated.

Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: An integration of research-based knowledge*. Charlotte, N.C: Information Age Pub.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Los Angeles: SAGE Publications.

CSS (2019). Annual Report 2018/2019. Retrieved on April 20, 2020 from https://cssalberta.ca/Portals/CSS/Documents/Mission/Annual-Reports/CSS_Annual%20Report%202018_v3_web.pdf?ver=2019-09-26-105330-250

CSS (2020). Annual Report 2019/2020. Retrieved on November 29, 2020 from https://online.flippingbook.com/view/287657/

CSS (2020). Our Mission. Retrieved on March 10, 2020 from https://cssalberta.ca/Our-Mission

Danseco, E. (2013). The five Cs for innovating in evaluation capacity building: Lessons from the field. *The Canadian Journal of Program Evaluation, 28*(2), 107.

Despard, M. R. (2016). Strengthening evaluation in nonprofit human service organizations: Results of a capacity-building experiment. *Human Service Organizations: Management, Leadership & Governance, 40*(4), 352-368. doi:10.1080/23303131.2016.1140101

Ebrahim, A. (2016; 2005). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly, 34*(1), 56-87. doi:10.1177/0899764004269430

Elson, P. R. (2011). *High ideals and noble intentions: Voluntary sector-government relations in Canada*. Toronto: University of Toronto Press. doi:10.3138/9781442690097

Fernsler, T. (2009). Create a culture of learning. *Nonprofit World, 27*(6), 27.

Fetterman, D. M. (2001). *Foundations of Empowerment Evaluation*. Thousand Oaks, Calif: Sage.

Fetterman, D. (2015). Building evaluation capacity to engage in empowerment evaluation A case of organizational transformation. *Empowerment Evaluation: Knowledge and Tools for Self-Assessment, Evaluation Capacity Building, and Accountability* (Second ed., pp. 233). Thousand Oaks: SAGE Publications, Inc. doi:10.4135/9781483387079.n15

Fierro, L. A., & Christie, C. A. (2017). Evaluator and program manager perceptions of evaluation capacity and evaluation practice. *American Journal of Evaluation, 38*(3), 376-392. doi:10.1177/1098214016667581

Frey, B. (2018). The SAGE encyclopedia of educational research, measurement, and evaluation (Vols. 1-4). Thousand Oaks,, CA: SAGE Publications, Inc. doi: 10.4135/9781506326139

Gagnon, F., Aubry, T., Cousins, J. B., Goh, S. C., & Elliott, C. (2018). Validation of the evaluation capacity in organizations questionnaire. *Evaluation and Program Planning, 68*, 166-175. doi:10.1016/j.evalprogplan.2018.01.002

Gerring, J., (2007;2006;). Case Study Research: Principles and practices. Cambridge.

Gill, S. J., 1947. (2010; 2013). *Developing a learning culture in nonprofit organizations*. Los Angeles: SAGE.

Guishard, M. (2008; 2009). The false paths, the endless labors, the turns now this way and now that: Participatory action research, mutual vulnerability, and the politics of inquiry. *The Urban Review,41*(1), 85-105. doi:10.1007/s11256-008-0096-8

Harper, G., & Suarez-Balcazar, Y. (2014). *Empowerment and participatory evaluation of community interventions: Multiple benefits* Taylor and Francis. doi:10.4324/9781315808703

Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation, 29*(3), 358-368. doi:10.1177/1098214008318798

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20. doi:10.1177/1525822X05282260

Jacob, S., Speer, S., & Furubo, J. (2015). The institutionalization of evaluation matters: Updating the international atlas of evaluation 10 years later. *Evaluation, 21*(1), 6-31. doi:10.1177/1356389014564248

Janzen, R., Ochocka, J., Turner, L., Cook, T., Franklin, M., & Deichert, D. (2017). Building a community-based culture of evaluation. *Evaluation and Program Planning, 65*, 163-170. doi:10.1016/j.evalprogplan.2017.08.014

Knutsen, W. L. (2017). Retaining the benefits of Government–Nonprofit contracting relationship: Opposites attract or clash? *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations,28*(4), 1373-1398. doi:10.1007/s11266-017-9874-9

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation, 33*(3), 307-338. doi:10.1177/1098214011434608

Laforest, R., & Smith, S. R. (2017). Nonprofits in a time of turbulence: Challenges and opportunities. *Nonprofit Policy Forum, 8*(2), 117-120. doi:10.1515/npf-2017-0021

Liket, K. C., Rey-Garcia, M., & Maas, K. E. H. (2014). Why Aren't evaluations working and what to do about it: A framework for negotiating meaningful evaluation in nonprofits. *American Journal of Evaluation, 35*(2), 171-188. doi:10.1177/1098214013517736

Lindeman, P. T., Bettin, E., Beach, L. B., Adames, C. N., Johnson, A. K., Kern, D., . . . Phillips, G. (2018). Evaluation capacity building—Results and reflections across two years of a multisite empowerment evaluation in an HIV prevention context. *Evaluation and Program Planning, 71*, 83-88. doi:10.1016/j.evalprogplan.2018.09.001

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage Publications.

Loud, M., and Mayne, J. (2014). Insights from Internal Evaluation Units. *Issues in Enhancing Evaluation Use.* Thousand Oaks: SAGE Publications, Inc. doi:10.4135/9781506335124.n1

Love, A. J. (1991). *Internal evaluation: Building organizations from within*. Newbury Park: Sage Publications.

Lu, S. K., Elliot, S. J., & Perlman, C. M. (2019). Perceived facilitators and barriers to evaluative thinking in a small, development NGO. *Canadian Journal of Program Evaluation, 34*(1) doi:10.3138/cjpe.43118

Mathison, S. (2005). *Encyclopedia of evaluation*. Thousand Oaks, Calif: Sage.

Mayne, J. (2009). Building an evaluative culture: The key to effective evaluation and results management. *The Canadian Journal of Program Evaluation, 24*(2), 1.

Mitchell, G. E., & Berlan, D. (2016). Evaluation and evaluative rigor in the nonprofit sector. *Nonprofit Management and Leadership, 27*(2), 237-250. doi:10.1002/nml.21236

Mitchell, G. E., & Berlan, D. (2018). Evaluation in nonprofit organizations: An empirical analysis. *Public Performance & Management Review, 41*(2), 415-437. doi:10.1080/15309576.2017.1400985

Muhammad, M., Wallerstein, N., Sussman, A. L., Avila, M., Belone, L., & Duran, B. (2014; 2015). Reflections on researcher identity and power: The impact of positionality on community based participatory research (CBPR) processes and outcomes. *Critical Sociology, 41*(7-8), 1045-1063. doi:10.1177/0896920513516025

MyM and E (2012, July 9). Exploring effective strategies for facilitating evaluation capacity building. Retrieved on March 19, 2020 from https://www.youtube.com/watch?v=uW-D9K0Ohl0

Najam, A. (2000). The four C's of government third Sector-Government relations. *Nonprofit Management and Leadership, 10*(4), 375-396. doi:10.1002/nml.10403

Parsons, B. L., Lovato, C. Y., Hutchinson, K., & Wilson, D. (2016). Building evaluation capacity through CLIPs: Communities of learning, inquiry, and practice. *New Directions for Evaluation, 2016*(151), 21-36. doi:10.1002/ev.20199

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health Services Research, 34(5 Pt 2), 1189-1208.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guilford Press.

Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Los Angeles, Calif: SAGE.

Pejsa, L. J. (2012). *Improving evaluation in the nonprofit sector: The promise of evaluation capacity building for nonprofit social service organizations in an age of accountability*

Available from Sociological Abstracts. (1081868659; 201239370). Retrieved from http://search.proquest.com.ezproxy.library.uvic.ca/docview/1081868659?accountid=14846

Perez Jolles, M., Collins-Camargo, C., McBeath, B., Bunger, A. C., & Chuang, E. (2017). Managerial strategies to influence frontline worker understanding of performance measures in nonprofit child welfare agencies. *Nonprofit and Voluntary Sector Quarterly, 46*(6), 1166-1188. doi:10.1177/0899764017728366

Poister, T. H., Hall, J. L., & Aristigueta, M. P. (2015; 2014). *Managing and measuring performance in public and nonprofit organizations: An integrated approach* (Second ed.). San Francisco, CA: Jossey-Bass & Pffeifer Imprints, Wiley.

Praznik, J. and Shields, J. (2018). An Anatomy of Settlement Services in Canada: A Guide. Building Migrant Resilience in Cities. Accessed on October 29, 2020 from https://bmrc-irmu.info.yorku.ca/files/2018/07/An-Anatomy-of-Settlement-Services-in-Canada_BMRCIRMU.pdf?x82641

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation, 29*(4), 443-459. doi:10.1177/1098214008324182

Preskill, H., & Boyle, S. (2008). Insights into evaluation capacity building: Motivations, strategies, outcomes, and lessons learned. *The Canadian Journal of Program Evaluation, 23*(3), 147.

Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation, 35*(1), 116-119. doi:10.1177/1098214013499439

Quarter, J., Mook, L., & Armstrong, A., 1951. (2009). *Understanding the social economy: A Canadian perspective*. Toronto; Buffalo: University of Toronto Press.

Renz, D. O., & Herman, R. D. (2016). *The Jossey-Bass Handbook of Nonprofit Leadership and Management*. Hoboken, New Jersey: Jossey-Bass & Pfeiffer Imprints, Wiley.

Rogers, A., Kelly, L. M., & McCoy, A. (2019). Evaluation literacy: Perspectives of internal evaluators in non-government organizations. *Canadian Journal of Program Evaluation, 34*(1) doi:10.3138/cjpe.42190

Rogers, A., Radcliffe, D., Babyack, S., & Layton, T. (2018). Demonstrating the value of community development: An inclusive evaluation capacity building approach in a nonprofit aboriginal and torres strait islander organisation. *Evaluation Journal of Australasia, 18*(4), 234-255. doi:10.1177/1035719X18803718

Runnels, V., Andrew, C., & Rae, J. (2017). Building evaluation capacity in a community-level program: Lessons learned from evaluating youth futures. *Canadian Journal of Program Evaluation, 32*(1) doi:10.3138/cjpe.30975

Saldaña, J., & Ebooks Corporation. (2011). *Fundamentals of qualitative research*. New York: Oxford University Press.

Senge, P. M. (1994). *The fifth discipline fieldbook: Strategies and tools for building a learning organization*. New York: Currency, Doubleday.

Senthanar, S., MacEachen, E., Premji, S., & Bigelow, P. (2020). "Can someone help me?" refugee Women's experiences of using settlement agencies to find work in canada. *Journal of International Migration and Integration, 21*(1), 273-294. doi:10.1007/s12134-019-00729-1

Shapiro, S. J., & Oystrick, V. (2018). Three steps towards sustainability: Spreadsheets as a data collection analysis system for non-profit organizations. *Canadian Journal of Program Evaluation, 33*(2) doi:10.3138/cjpe.31157

Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process.*American Journal of Evaluation, 37*(2), 193-215. doi:10.1177/1098214015615230

Stitt-Bergh, M. (2016). Assessment capacity building at a research university. *New Directions for Evaluation, 2016*(151), 69-83. doi:10.1002/ev.20196

Stoecker, R. (2007). The research practices and needs of non-profit organizations in an urban center. *Journal of Sociology and Social Welfare, 34*(4), 97-119.

Suarez-Balcazar, Y., & Harper, G. W. (2003; 2004; 2014). *Empowerment and participatory evaluation of community interventions: Multiple benefits*. New York: Haworth Press. doi:10.4324/9781315808703

Suarez-Balcazar, Y., Taylor-Ritzler, T. & Morales-Curtin, G. (2015). Building evaluation capacity to engage in empowerment evaluation: a case of organizational transformation. In Empowerment evaluation (pp. 233-258). SAGE Publications, Inc, https://www-doi-org.ezproxy.library.uvic.ca/10.4135/9781483387079

Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D. B., & Balcazar, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation, 34*(2), 190-206. doi:10.1177/1098214012471421

Treasury Board Secretariat (2016a). Policy on Results. Government of Canada. Retrieved on March 9, 2020 from https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300

Umar, S., & Hassan, S. (2019). Encouraging the collection of performance data in nonprofit organizations: The importance of organizational support for learning. *Public Performance & Management Review,42*(5), 1062-1084. doi:10.1080/15309576.2018.1481118

van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review, 25*(3), 267-281. doi:10.1080/15309576.2002.11643661

Whitaker, G.P., Altman-Sauer, L., and Henderson, M. (2004). Mutual Accountability between Governments and Nonprofits: Moving Beyond “Surveillance” to “Service”.

Wholey, J. S., Hatry, H. P., Newcomer, K. E., 1949, & Ebooks Corporation. (2004; 2003). *Handbook of practical program evaluation* (2nd ed.). San Francisco: Jossey-Bass.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Sage.

Zarinpoush, F., Canadian Electronic Library (Firm), & Imagine Canada. (2006). *Project evaluation guide for nonprofit organizations: Fundamental methods and steps for conducting project evaluation*. Toronto, Ont: Imagine Canada.

Appendices

Appendix 1. Items on the Evaluation Capacity Assessment Instrument, adapted to context of Immigration and Settlement Service.

All items in Section 1 - 3 are Likert statements that respondents rate on a scale of 1 - 4: 1, strongly disagree; 2, somewhat disagree; 3, somewhat agree; and, 4, strongly agree.

Demographic Questions

1. What title best describes your current position?
    a. Program Manager / Senior Leadership
    b. Team Leader
    c. Program / Project / Volunteer Coordinator
    d. Settlement Counsellor
    e. Language Assessor
    f. Program Assistant
    g. Other Support Staff
2. How many years have you worked for Catholic Social Services?
    a. Less than 1 year
    b. 2 - 5 years
    c. 5 - 10 years
    d. 10 - 15 years
    e. More than 16 years

Section 1: Individual Factors

Awareness: I think that an evaluation will…

1. Will help me to better understand our programs.
2. Will inform changes to my service delivery practices with clients.
3. Will justify future funding for our programs.
4. Will help to convince managers that changes are needed in my/our programs.
5. Is absolutely necessary to improve our programs.
6. Should involve our clients in the evaluation process.
7. Will influence government policy relevant to our programs.
8. Is unnecessary because we already know what is best for our clients.
9. Is too complex for frontline staff to do.

Motivation: I am motivated to…

1. Learn about program evaluation and how it impacts my work.
2. Support the evaluation of our programs.
3. Participate in training on the topic of evaluation.
4. Have conversations with my colleagues about why evaluation is necessary.

Competency: I know how to…

1. Develop an evaluation plan.

2. Clearly state measurable goals and objectives for my/our program.
3. Identify strategies to collect information from our clients.
4. Define the intended outcomes of my/our program.
5. Decide what questions to answer in a program evaluation.
6. Decide from whom to collect the information.
7. Analyze evaluation information.
8. Develop recommendations based on evaluation results.
9. Write an evaluation report.
10. Evaluate a program without support from others.
11. Conduct an evaluation of my/our program/s with support from others.
12. Present evaluation findings in a team meeting.

SECTION 2: ORGANIZATIONAL FACTORS

LEADERSHIP

1. Program managers provide effective leadership.
2. Staff understands how everyone's duties fit together as part of the overall mission of the Service.
3. Program managers communicate program goals and objectives clearly.
4. Program managers share funder requirements and expectations openly with staff.
5. Program managers share documents, such as program logic models, openly with staff.
6. Program managers have a clear plan for accomplishing goals.
7. Program managers have realistic expectations of what staff can accomplish given the resources they have available.

LEARNING CLIMATE: THE SERVICE I WORK IN FOSTERS AN ENVIRONMENT IN WHICH…

1. Evaluation information is shared in team and all-staff meetings.
2. Staff are supported to introduce new approaches to service delivery or other aspects of our work.
3. It is easy for staff to meet regularly and discuss new ideas.
4. Staff are provided opportunities to assess how well they are doing, what they can do better, and what is working.
5. Staff can encourage managers and peers to make use of evaluation findings.
6. Staff respects each other's perspectives and opinions.
7. Staff errors lead to teachable moments rather than criticisms.
8. Staff participate in making long-term plans for their program.
9. Staff concerns are ignored in most decisions regarding strategic planning and evaluation.

RESOURCES: IN OUR PROGRAMS…

1. Translation and interpretation is provided to clients to support their participation in evaluation.
2. Staff has time to conduct evaluation activities (ie. provide input on survey/focus group questions, distribute a survey, invite clients to focus groups, conduct phone interviews).

3. Staff has access to adequate technology to input information into computerized records (Sharevision, spreadsheets, other documents or databases).

RESOURCES (MANAGEMENT ONLY): IN MY PROGRAMS…

1. Staff have access to technology to produce reports or summaries of their activities and client information (ie. computerized databases).
2. Resources are allocated to staff training, including evaluation capacity training.
3. Funders provide resources to conduct evaluation.
4. Funders provide leadership for conducting evaluation, such as what is considered a valid/defensible evaluation.
5. Service leadership engages in ongoing dialogue with funders regarding evaluation.

SECTION 3: EVALUATION CAPACITY OUTCOMES

MAINSTREAMING: EVALUATION AS PART OF YOUR JOB.

1. Our programs gather information from diverse stakeholders to determine how well programs are doing.
2. Our programs have adequate records of past program evaluations and the changes that were implemented as a result.
3. I have access to the information I need to make decisions regarding my work.
4. I am able to integrate evaluation activities into my daily work practices.

USE OF FINDINGS: PLEASE INDICATE THE EXTENT IN WHICH YOUR PROGRAM CURRENTLY USES EVALUATION FOR THE FOLLOWING PURPOSES.

1. To report to a funder.
2. To improve service or programs.
3. To get additional funding.
4. To assess the implementation strategy of a program.
5. To assess the quality of our programs.
6. To improve community outreach.
7. To make informed decisions.
8. To train staff.
9. To develop best practices.
10. To eliminate unneeded services or programs.

Appendix 2. ECAI Data at ISS.

Table shows total responses per statement, mean and mode per statement. The highest scores have been highlighted in green and the lowest in red.

	EC Sub-Factor	Question	Responses (n)	Mean	Mode	Sub-Factor Mean
Individual Factors	**Awareness** *I think that an evaluation…*	Will help me to better understand our programs.	41	3.66	4.00	3.31
		Will inform changes to my service delivery practices with clients.	41	3.68	4.00	
		Will justify future funding for our programs.	41	3.63	4.00	
		Will help to convince managers that changes are needed in my/our programs.	41	3.61	4.00	
		Is absolutely necessary to improve our programs.	41	3.49	4.00	
		Should involve our clients in the evaluation process.	41	3.59	4.00	
		Will influence government policy relevant to our programs.	41	3.29	3.00	
		Is unnecessary because we already know what is best for our clients.	41	2.49	1.00	
		Is too complex for frontline staff to do.	41	2.39	1.00	
	Motivation *I am motivated to…*	Learn about program evaluation and how it impacts my work.	41	3.61	4.00	3.59
		Support the evaluation of our programs.	41	3.63	4.00	
		Participate in training on the topic of evaluation.	41	3.66	4.00	
		Have conversations with my colleagues about why evaluation is necessary.	40	3.45	4.00	
	Competence *I know how to…*	Develop an evaluation plan.	41	2.78	3.00	3.06
		Clearly state measurable goals and objectives for my/our program.	41	3.15	3.00	
		Identify strategies to collect information from our clients.	41	3.27	3.00	
		Define the intended outcomes of my/our program.	41	3.32	3.00	
		Decide what questions to answer in a program evaluation.	41	3.05	3.00	
		Decide from whom to collect the information.	41	3.27	3.00	
		Analyze evaluation information.	41	3.17	3.00	
		Develop recommendations based on evaluation results.	41	3.37	3.00	
		Write an evaluation report.	40	2.80	3.00	
		Evaluate a program without support from others.	41	2.02	2.00	

		Conduct an evaluation of my/our program/s with support from others.	41	3.17	4.00	
		Present evaluation findings in a team meeting.	41	3.32	4.00	
Organizational Factors	**Leadership**	Program managers provide effective leadership.	39	3.33	4.00	**3.05**
		Staff understands how everyone's duties fit together as part of the overall mission of the Service.	40	2.90	3.00	
		Program managers communicate program goals and objectives clearly.	40	3.08	3.00	
		Program managers share funder requirements and expectations openly with staff.	40	3.18	4.00	
		Program managers share documents, such as program logic models, openly with staff.	40	2.75	3.00	
		Program managers have a clear plan for accomplishing goals.	40	3.05	3.00	
		Program managers have realistic expectations of what staff can accomplish given the resources they have available.	40	3.10	4.00	
	Learning Climate *The Service I work in fosters an environment in which…*	Evaluation information is shared in team and all-staff meetings.	39	3.03	3.00	**2.91**
		Staff are supported to introduce new approaches to service delivery or other aspects of our work.	40	3.10	4.00	
		It is easy for staff to meet regularly and discuss new ideas.	40	2.95	3.00	
		Staff are provided opportunities to assess how well they are doing, what they can do better, and what is working.	39	3.00	3.00	
		Staff can encourage managers and peers to make use of evaluation findings.	40	2.90	3.00	
		Staff respects each other's perspectives and opinions.	40	3.30	3.00	
		Staff errors lead to teachable moments rather than criticisms.	40	3.20	3.00	
		Staff participate in making long-term plans for their program.	40	2.75	3.00	
		Staff concerns are ignored in most decisions regarding strategic planning and evaluation.	40	1.98	2.00	
	Resources *In our programs…*	Translation and interpretation is provided to clients to support their participation in evaluation.	40	3.43	4.00	**2.98**
		Staff has time to conduct evaluation activities (ie. provide input on survey/focus group questions, distribute a survey, invite clients to focus groups, conduct phone interviews).	39	2.97	4.00	
		Staff has access to adequate technology to input information into computerized records (Sharevision, spreadsheets, other documents or databases).	40	3.20	4.00	

	Resources (Management Only) *In my programs…*	Staff have access to technology to produce reports or summaries of their activities and client information (ie. computerized databases).	15	3.20	3.00	
		Resources are allocated to staff training, including evaluation capacity training.	15	2.87	3.00	
		Funders provide resources to conduct evaluation.	15	2.73	3.00	
		Funders provide leadership for conducting evaluation, such as what is considered a valid/defensible evaluation.	13	2.38	2.00	
		Service leadership engages in ongoing dialogue with funders regarding evaluation.	15	3.07	3.00	
Evaluation Capacity Outcomes	**Mainstreaming** *Evaluation as part of your job.*	Our programs gather information from diverse stakeholders to determine how well programs are doing.	37	3.19	3.00	**3.09**
		Our programs have adequate records of past program evaluations and the changes that were implemented as a result.	36	2.92	3.00	
		I have access to the information I need to make decisions regarding my work.	37	3.22	4.00	
		I am able to integrate evaluation activities into my daily work practices.	36	3.06	3.00	
	Use of Findings *Please indicate the extent in which your program currently uses evaluation for the following purposes.*	To report to a funder.	35	3.17	3.00	**2.88**
		To improve service or programs.	36	2.97	3.00	
		To get additional funding.	35	2.94	3.00	
		To assess the implementation strategy of a program.	37	2.86	3.00	
		To assess the quality of our programs.	36	3.00	3.00	
		To improve community outreach.	37	2.73	3.00	
		To make informed decisions.	36	3.00	3.00	
		To train staff.	37	2.68	3.00	
		To develop best practices.	36	2.97	3.00	
		To eliminate unneeded services or programs.	37	2.43	2.00	

Appendix 3. Sample Slide from Presentation of the ECAI results at ISS.

Summary - Individual Factors

Highest Rated

Awareness
- Will help me to better understand our programs. (3.66)
- Will inform changes to my service delivery practices with clients. (3.68)

Motivation
- Participate in training on the topic of evaluation. (3.66)

Competence
- Develop recommendations based on evaluation results. (3.37)

Lowest Rated

Awareness
- Is unnecessary because we already know what is best for our clients. (2.49)
- Is too complex for frontline staff to do. (2.39)

Motivation
- Have conversations with my colleagues about why evaluation is necessary. (3.45)

Competence
- Develop an evaluation plan. (2.78)
- Evaluate a program without support from others. (2.02)

- Are the results similar to your experience?