Developing the Program Evaluation Framework for Investment Agriculture Foundation of British Columbia

By

Zulfiqar Ali
B.Sc. (Hons.), Lahore University of Management Sciences, Pakistan, 2013
M.A., International Christian University, Japan, 2016

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF PUBLIC ADMINISTRATION

in the School of Public Administration

Supervisory Committee

Developing the Program Evaluation Framework for Investment Agriculture Foundation of British Columbia

Zulfiqar Ali
B.Sc. (Hons.), Lahore University of Management Sciences, Pakistan, 2013
M.A., International Christian University, Japan, 2016

Supervisor: Dr. Kimberly Speers, Assistant Teaching Professor and Graduate Advisor, University of Victoria

Committee member: Dr. Robert Newell, Canada Research Chair in Climate Change, Biodiversity and Sustainability, Royal Roads University

Abstract

This thesis aimed to gain an in-depth understanding of effective program evaluation frameworks, particularly in the agriculture sector. Specifically, the analysis was focused on assessing monitoring and evaluation of public and private sector projects to assist the British Columbia Investment Agriculture Foundation (IAF), the client for this thesis, in identifying smart practices as a way to support constant improvement in their organization. This research involved conducting a literature review of the most recent and relevant literature on program evaluation, particularly works that related to nonprofit organizations in the agriculture sector, interviewing IAF staff, developing a jurisdictional scan of program evaluation frameworks in the Netherlands and New Zealand, and conducting a review of existing IAF evaluation documents to identify and discuss key themes for an effective program evaluation framework and provide examples of smart evaluation practices that may be adapted by IAF. The recommendations include integrating formative and summative evaluation practices, developing targeted programs with well-defined key performance indicators (where possible), and capitalizing on data visualization software for monitoring and reporting on project goals in real-time.

Key words: Program evaluation, agri-sector program evaluation, evaluation frameworks, smart evaluation practices

Table of Contents

Abstract iii
Table of Contents iv
Acknowledgements vii
Dedication viii
Chapter 1: Introduction 1
Issue Statement and Significance of the Topic 2
Main Research Question 2
Positionality Statement 3
Structure of the Thesis 4
Chapter 2: Literature Review 5
Introduction 5
Defining Program Evaluation 5
Evaluation and Accountability 6
Monitoring and Evaluation 6
Evaluation and Performance Measurement 7
Program Evaluation in Nonprofits 7
Program Evaluation for Agricultural Programs 8
Summary: Literature Review 11
Theoretical Framework 12
Chapter 3: Methodology and Methods 14
Methodology 14
Methods 15
Key Informant Interviews 15
Document Review 16
Data Analysis 16
Limitations and Strengths 16
Limitations 16
Strengths 17

Chapter 4: Current State Analysis - Interview Results 18
Theme 1 - Background Information 18
Theme 2 - Data Collection and Developing the Baseline 18
Theme 3 - Current State – Performance Indicators 19
Theme 4 - Logic Models and Evaluation Frameworks 21
Theme 5 - Evaluation Success and Client Satisfaction 21
Theme 6 - Strengths and Challenges – IAF Program Evaluation 22
Theme 7 – IAF and Program Evaluation 23
Summary of Interview Findings 24
Chapter 6: Current State Analysis - IAF Document Review Findings 26
Financial and Project Reporting Templates 26
Logic Model Templates and Frameworks 29
Summary of the IAF Document Review chapter 35
Chapter 5: Jurisdictional Scan Findings 36
The Netherlands: 36
Results 36
New Zealand: 39
Results: 39
Summary of Jurisdiction Scan Findings 44
Chapter 7: Discussion and Analysis 46
Main Themes in Research 46
Theme 1 - IAF evaluations tend to be accountability-focused 46
Theme 2 - IAF needs to conduct more field visits to monitor and measure program impacts 46
Theme 3 - IAF program goals need to be more targeted 47
Theme 4 - IAF needs to integrate formative evaluation components in their evaluation practices 48
Theme 5 - IAF should integrate data visualization software in their evaluation practices 48
Chapter 8: Conclusions and Recommendations 50
Recommendations 51
Moving Forward 52

References 53
Appendices 59
Appendix A: Interview Invite Templates 59
Appendix B: Consent Form Template 61
Appendix C: Interview Questionnaire 65

Acknowledgements

I would like to express my sincerest gratitude to my research supervisor, Dr. Kimberley Speers who has provided me constant guidance and support over the last year in developing, designing, and reviewing thesis proposals and drafts, and in getting the Human Research Ethics Board approval for this research. It would have been impossible to complete this thesis in time for graduation in early Fall 2022 without her.

I am also very thankful to Dr. Robert Newell for always providing poignant and timely feedback on my research proposals and thesis drafts. Sincere gratitude is also owed to Natalie Janssens, Program Manager at IAF, for bringing this research topic to us, helping coordinate IAF staff interviews for this research project and always providing timely feedback on all research proposal and thesis drafts. I also appreciate all the interviewees for participating in this research. I would also like to thank Dr. John Volpe for agreeing to be an external evaluator for this thesis.

Finally, I am grateful for the support provided by the administrative staff in the MPA program at the University of Victoria in helping me keep up with all the program requirements.

Dedication

This thesis is dedicated to my family and friends, who have kept me going in the last two years.

Chapter 1: Introduction

The purpose of this thesis is to identify effective program evaluation frameworks for monitoring and evaluating the effectiveness of public and private sector projects in the agricultural sector with the intention of applying them to the Investment Agriculture Foundation (IAF) in British Columbia (BC) if not already in place and adopting frameworks relevant to the goals of IAF. The emphasis of this thesis is conducting research on the program evaluation of marketing and environmental programs as those make up the majority of IAF's current program contracts with the government, with a broader focus on the British Columbia agriculture, agri-food, and agri-tech sectors.

The Investment Agriculture Foundation is acting as an informal client for this thesis to provide a pragmatic approach to research, which is consistent with the discipline of public administration. The Foundation's mandate is to be an industry-led, nonprofit society that creates funding opportunities to support a thriving BC agriculture and agri-food sector through the effective delivery of programs (Investment Agriculture Foundation, 2021). Over the last 25 years, IAF has delivered over $200 million worth of programs to the industry on behalf of the federal and provincial governments for the agriculture and agri-food sector in British Columbia.

In delivering these programs and providing funding, the Investment Agriculture Foundation's current program evaluation process largely consists of summarizing the amount of funding delivered to stakeholders and collecting program-specific reporting metrics as required by the funders. Given the challenges of conducting impact and outcome evaluations and having the evaluation framework already decided for most programs by a funding partner at the outset of a project, few evaluations have assessed their programs' effectiveness and impacts in IAF and in nonprofit organizations in general. This research project aims to develop evaluation frameworks to help IAF measure outcomes and impacts to provide greater value to the funder as well as working within the limitations of the funding and evaluation system.

The research involves a jurisdictional review of program evaluations in other agri-sector dominant economies such as New Zealand and the Netherlands to identify any lessons learned or feasible and relevant evaluation frameworks. The study also includes key informant interviews conducted with relevant staff members at IAF and examines the evaluation techniques used by both the public and private sector to provide a smart practices' guide for IAF's consideration in implementing a program evaluation framework. In doing so, the thesis is intended to contribute to general knowledge on effective evaluation practices and frameworks for programs that fund and support agriculture, agri-food, and agri-tech initiatives, and how these practices/frameworks can be adapted to different geographic and organizational contexts.

Issue Statement and Significance of the Topic

The issue being addressed in this thesis is administrative and procedural in nature. Currently, IAF's program evaluation activities are varied and primarily based on a funding agency's requirements for each program that the organization is contracted to deliver. According to the client, the evaluations consist of a summary of funding delivered to stakeholders as opposed to an assessment of the program's efficacy and impact.

The Foundation's main funding partners (provincial and federal government agencies) have expressed a desire to demonstrate the value for money and the impact of programs funded through taxpayer dollars as well as those areas requested by the stakeholders. Yet there are challenges to doing so given the way the funding system is currently structured in which program objectives for most programs are predetermined by the funders and IAF acts more as a delivery agent/implementing partner for those programs. Therefore, IAF's role in designing program goals or objectives for these programs is severely constrained.

Despite these constraints, this study still provides value to IAF in that it outlines smart practices for IAF to evaluate programs using tools such as logic models, specific activity reporting metrics, and surveys to help better track and measure inputs, activities, outputs, outcomes, and impacts, irrespective of who sets the program objectives. Moreover, this research will also be useful for IAF going forward as they play a more developmental role in some programs, which may involve developing program objectives or performance indicators collaboratively with the funding agencies. In this case, the thesis and its learnings can be presented in such a way (wherever possible) that it could be used by IAF as a "key service offered" when bidding to deliver new government programs.

Main Research Question

The main research question this thesis explored is: *What makes for an effective program evaluation framework(s) for IAF?*

In this thesis, program effectiveness refers to the degree to which a program is deemed useful in meeting its intended goals and objectives under more real-world conditions (Chen, 2005).

The research also explores the following secondary research questions:

- What is program evaluation?
- What are the different evaluation types?
- What are the key elements of an evaluation?
- What are the program evaluation challenges faced by nonprofits?

- What are some program evaluation challenges faced by nonprofits in the agriculture and/or agri-tech, agri-food sectors?
- What are some of the strategies to cope with challenges faced by nonprofits in program evaluation?

While many of these concepts will be further discussed in the literature review, to make it easier to follow and understand the thesis at the outset, key terms related to agriculture are defined below:

Agriculture: Agriculture is the science of cultivating the soil, growing crops, and raising livestock. It includes the preparation of plant and animal products for people to use and their distribution to markets (National Geographic, n.d.).

Agri-tech: Agri-tech refers to the fusion of innovation and technology applied to the agriculture, food processing, and seafood sectors. (Government of British Columbia, n.d.).

Agri-food: The agri-food sector refers to a wide range of industries, from primary agriculture and aquaculture to food, seafood, and beverage processing (Agri-Food Economic Strategy Table, 2017).

Positionality Statement

I became interested in agri-sector policy evaluation during my two-year tenure as the Research Associate at Pakistan Agricultural Coalition in Karachi. There, I helped incubate and pilot a farm-machinery services' business for farmers to provide them modern farm machinery needed for improving farm productivity. A key part of the pilot was to develop a monitoring and evaluation framework, collect and analyze field data to identify challenges, and document progress for investors through interim and final project reports. More recently, in my first co-op as an Analyst, Public Sector Consulting at MNP, I was involved in collating key findings from 150+ individual & focus group interviews to evaluate primary, continuing care and acute care services in Alberta. In addition, I reviewed stakeholder analysis to address the clinical care challenges faced by the Gitsegukla First Nation Community. As such, the purpose of this thesis is to gain an in-depth understanding of program evaluation frameworks, particularly in the agri-sector, for monitoring and evaluation of public and private sector projects.

Therefore, some of my personal, professional and/or intellectual positionalities (identities, contexts, experiences, and perspectives) may have inspired the selection of this research theme but they do not necessarily adversely affect its outcome since I can be considered as an outsider (international student) and someone with no vested interest in the outcome of this research. It will be important, however, that I, conscientiously, or not, do not reify, disrupt, and/or change the constructs of my positionality through this research process and follow the research process with integrity and objectivity.

Structure of the Thesis

The following literature review chapter introduces, defines, and discusses key themes such as program evaluation, the relationship between evaluation and accountability, and monitoring and evaluation, and the similarities and differences between evaluation and performance measurement. The section also provides the most recent and relevant literature on the challenges of conducting program evaluations in the nonprofit sector, particularly nonprofits that work in the agriculture sector and provides examples of some evaluation practices of modern agriculture sector nonprofits that may be adapted at IAF. Finally, the chapter also introduces the program theory theoretical framework for this research.

The methodology and methods chapter introduces the primary research methodology, which is a current state analysis of program evaluation and a jurisdiction scan. The methodology is qualitative in focus and the methods used are key information interviews with IAF staff, document review of existing IAF evaluation and reporting templates, and data analysis of these findings. The chapter also discusses the limitations and strengths of the research.

The findings section of the thesis has three chapters – jurisdictional scan and two current state analysis chapters consisting of interviews and a document review of IAF reporting templates and documents.

The discussion and analysis chapter analyzes, discusses, and synthesizes findings from the interviews, jurisdictional scan and the IAF document review. This chapter also identifies the strategic implications of the results and areas of further research.

The final chapter provides concluding thoughts and highlights recommendations for IAF for improving their current program evaluation framework. It also shares the lessons learned and limitations of the study and highlights the future scope of research.

Chapter 2: Literature Review

Introduction

The thesis uses a narrative literature review, which is defined as "a comprehensive, critical and objective analysis of the current knowledge on a topic" (CSU Library, 2021). A narrative literature review forms an essential part of the research process and helps establish a theoretical framework and context for the research (CSU Library, 2021). The main themes explored in the literature review are:

- Program evaluation
- Key themes in program evaluation
    - Accountability, performance measurement & monitoring, process, outcome, impact, and performance monitoring,
- Program evaluation in nonprofit organizations
- Program evaluation in agricultural programs
- Program theory framework

The literature review assesses independent empirical and qualitative studies from online databases such as Jstor, Sage, Taylor & Francis, and Elsevier among others. The main search terms used were: "agriculture," "program evaluation framework", "agriculture programs evaluation", "program theory evaluation framework", "logic models", "smart practices," "agriculture-sector logic models", "agri-sector evaluation framework", "agri-food evaluation framework" and "agri-tech evaluation framework," The search aimed to find most relevant academic studies on the topic published within the last twenty years.

Defining Program Evaluation

While there are many definitions of program evaluation, at its core, it can be understood as the "process of determining the merit, worth, and value of things" (Verdung 1997, p. 2). Verdung further notes that it is often a retrospective scrutiny of the administration, inputs, outcomes, and challenges faced in the implementation of a policy or program (Verdung 1997, p. 1); however, it can also be a proactive endeavor to develop a program or policy, anticipate and rectify its potential challenges and to measure and record key performance indicators of a policy or program. The former, often called summative or outcome evaluations, are conducted towards the end of the program to provide decisive input about the success or failure of the program activities vis-à-vis program objectives (Alkin & Vo, 2018, p. 13).

On the other hand, formative or process evaluations are often conducted in the initial stages of a program to identify any challenges and determine course-correction (Alkin & Vo, 2018, p. 13). Yet the difference between the two is not absolute; for instance, determining whether a policy has had an impact often involves asking questions about how it has done so, for whom, why, and

under what conditions (Fox, Grimm and Caldiera, 2016). A third most common form of evaluation are impact evaluations that focus on long-term program goals and issues of causality (OVC, 2010). The difference between the three types of evaluation is not always as described above and can vary in different situations and context. Therefore, choosing a specific evaluation type depends on asking question about program objectives such as what objectives the program has achieved, how it has done so, for whom, why, and under what conditions (Fox, Grimm and Caldiera, 2016).

This thesis uses Patton's definition of program evaluation as "the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding" (Patton, 2008, p. 39) as it encapsulates the desired features of program evaluation intended for this research. According to Peer (2017), an effective program evaluation framework allows organizations to engage with stakeholders, reflect on both failures and successes and learn from them to make evidence-based decisions.

Evaluation and Accountability

According to Carman (2009, p. 374) nonprofits, on average, are becoming more accountable than ever, and more focused on outcomes and results than in the past. Alkin and Klein (1972, p. 2) defined accountability from the perspective of evaluation as "a negotiated relationship in which the participants agree in advance to accept specified rewards and costs on the basis of evaluation findings as to the attainment of specified ends." The essence of this definition is the mutual (and negotiated) agreement between participants in advance about the criteria and the evaluation findings that will determine acceptability of results. The evaluation criteria are pre-specified, as are the rewards and penalties attached to compliance/non-compliance (Alkin and Klein, 1972, p. 2-3). In this context, the focus on accountability and performance evaluation in nonprofits comes from several factors such as the increasing number of nonprofits being involved in the delivery of public goods and services, prompting governments to ensure proper accountability of the funds dispensed. Another reason is that "highly visible" financial scandals involving prominent nonprofit organizations have called into question the integrity of nonprofit organizations (Carman 2009, p. 375).

Some scholars such as Salamon (2016) attribute the increased emphasis on accountability to the commercialization of the nonprofit sector and the competition between for-profits and nonprofits for funding and resources and consequently the need to be accountable for both. Accountability can also be a useful way for nonprofits to manifest, exert and maintain their independence vis-a-vis their mission (Carman 2009, p. 375).

Monitoring and Evaluation

Monitoring is a key element of the evaluation process. Irrespective of the evaluation type, monitoring helps ensure programs are implemented in accordance with their design and

objectives. It helps answer questions such as “Are we doing the right thing? Are we doing it right?” (Suvedi & Stoep 2016, p. 3). Thereby, monitoring tracks inputs, activities, outputs, outcomes and impacts throughout the project and yields valuable data for either course-correction or future improvements about the project. In this way, monitoring and evaluation are “synergistic.” (World Bank, n.d.)

Monitoring information is a necessary but is not a sufficient input to the conduct of rigorous evaluations. While monitoring information can be collected for ongoing management purposes, sole reliance on such information can introduce distortions as monitoring can typically cover only certain dimensions of a program's activities. In contrast, evaluation has the potential to provide a more balanced interpretation of performance (World Bank, n.d.).

Evaluation and Performance Measurement

Performance measurement is an intrinsic part of program evaluation as exemplified in the program evaluation literature. It “evaluates how well organisations are managed and the value they deliver for customers and other stakeholders” (Moullin, 2002, p. 188). Program evaluations assess whether the program is meeting its designated performance indicators using measurement data, which also helps identify areas of improvement and other barriers affecting program performance (United States Environmental Protection Agency, 2022).

As noted in the literature, there are some key differences between performance measurement and program evaluation. For instance, performance measurement is an ongoing process in which data are collected, analyzed, and reported close to real time as possible, giving staff immediate and actionable feedback, whereas an evaluation is discrete; carried out during particular periods of a program’s development or implementation and over a specified timeframe (Tatian, 2016).

Secondly, performance measurement often uses data collected routinely during program operations, such as clients’ use of services whereas an evaluation will often expand on those data sources, by collecting additional data through surveys, direct observation, or other means. Although not always, performance measurement is also mostly done by program staff, whereas evaluation is typically carried out by people outside the program to get a more objective assessment than might be provided by internal program staff or to access skills not available within the nonprofit (Tatian, 2016). Both performance measurement and program evaluation serve essential functions for performance management and improvement (Taylor & Liadsky, 2016).

Program Evaluation in Nonprofits

Nonprofit organizations are being increasingly called upon to develop and implement effective program evaluation frameworks to ensure efficiency in their services, which are often preconditions for receiving public and private sector funding (Carnochan, Samples, Myers, & Austin, 2013, p. 1). It is also an effective strategy for them to attract and retain a well-qualified

workforce, meet stakeholder expectations, and chart long-term sustainability (Carnochan, Samples, Myers, & Austin, 2013, p. 2). As Wing (2004) succinctly stated over 15 years ago, "two trends in philanthropy have been gaining momentum in recent years—an increased focus on measurable outcomes on one hand and greater investment in capacity building or organizational effectiveness on the other" (Carman 2009, p. 374).

The research on nonprofit sector program evaluation has increased steadily over the past two decades. Yet, it is dwarfed by the research on the private sector program evaluation, which because of its focus on profitability and shareholder value mostly, is of limited relevance to the nonprofit sector (Mitchell & Berlan, 2018, p. 416). For example, one main difference between for-profits and nonprofits is the complexity of performance measurement. Nonprofits often pursue outcomes that are often difficult both to measure and achieve (Lee & Nowell, 2015). Another argument can be made that employees in nonprofit organizations tend to be highly intrinsically motivated and identify with the mission of the organization compared to for-profit employees, who tend to be characterized as extrinsically motivated (Walk & Kennedy 2016, p. 5). From an organizational perspective therefore, most nonprofit organizations cannot easily reward employee performance with financial incentives such as salary raises or bonus payments given their organizational characteristics, thereby needing different performance management structures (Walk & Kennedy 2016, p. 5). Moreover, the research on performance measurement in the nonprofit sector also needs to be more integrated, comparable to the public and private sectors (Lee & Nowell, 2015).

To address these challenges, some scholars have highlighted three core strategies for effective performance measurement for nonprofits that involve (a) nurturing local stakeholder involvement, (b) creating specific goals logically linked to metrics that measure, and (c) continually fine-tuning measures and goals strategically linked to balancing the needs of funding partners (Carnochan, Samples, Myers, & Austin, 2013, p. 3)**.**

Program Evaluation for Agricultural Programs

Monitoring and evaluation (M&E) in agricultural research and development (R&D) is relatively recent, driven primarily by funder requirements, to meet accountability purposes (Stone-Jovicich, Percy, White, Turner, & Chen, 2019). This has led to calls for a broader array of monitoring, evaluation, and learning (MEL) design and planning processes in not only enhancing the impact of science contributions but more broadly in supporting research and innovation in Agriculture Innovation Systems (AIS) (Stone-Jovicich, Percy, White, Turner, & Chen, 2019).

Even more limited is the institutionalization and capacities of various agriculture-sector evaluation agencies, particularly in the developing countries (FAO, 2020). A 2020 (pre-pandemic) study of public sector agriculture programs in 23 developing countries (from Africa, Central and South Asia and Latin America) by Food and Agriculture Organization (FAO) that assessed their evaluation frameworks and practices concluded that evaluation capacities and

activities were “generally overlooked” (p. 26). It also found that evaluations were largely donor-driven, and therefore accountability-based, and that the program evaluations were “rare” for “national” (public sector) programs. Specifically, the study found four key challenges faced by the government agencies centered on agriculture in these countries: i) insufficient staff capacities, ii) lack of institutionalization and/or operationalization of monitoring and evaluation systems and tools, iii) budget constraints, and iv) unavailable or low-quality data (FAO 2020, p. 26).

Program evaluation research in the agriculture sector generally focuses on extension services, such as getting estimates of future prices for farm products, research on new products and knowledge about how to use particular inputs (e.g., the timing and intensity of fertilizer use) rather than program management and delivery (Anderson & Feder 2004, p. 42). Even then, one USAID study found most agriculture extension evaluation programs to be “complex” and “messy” and not focused on “long-term” goals (Suvedi & Stoep 2016, p. 5). This is often because of the lack of staff capacity and funds to conduct evaluation. Often, extension services’ evaluations focus on just collecting demographic information about participants or administering end-of-program evaluation surveys and then associating these perceptions with socio-demographic characteristics such as sex, race, residence, or educational level to report these to the national level decision makers. Although information about outputs and perceptions is helpful, it does not provide information about the impact of the extension program or service (Suvedi & Stoep 2016, p. 5). There is also little attention on systematic participation by the farming community in problem definition, problem solving, and extension programming (Anderson & Feder 2004, p. 47).

To address these challenges, in 2010, the G20 countries called for “innovative, results-based mechanisms… to achieve breakthrough innovations in food security and agriculture development… in poor countries” in the Toronto and Seoul Summits (G20 Summit 2010, Annex III). This led to the development of AgResults led by a core group of countries consisting of Australia, Canada, the United States, and the United Kingdom, and in partnership with the Bill and Melinda Gates Foundation, the World Bank, and the global development advisory firm Dalberg.

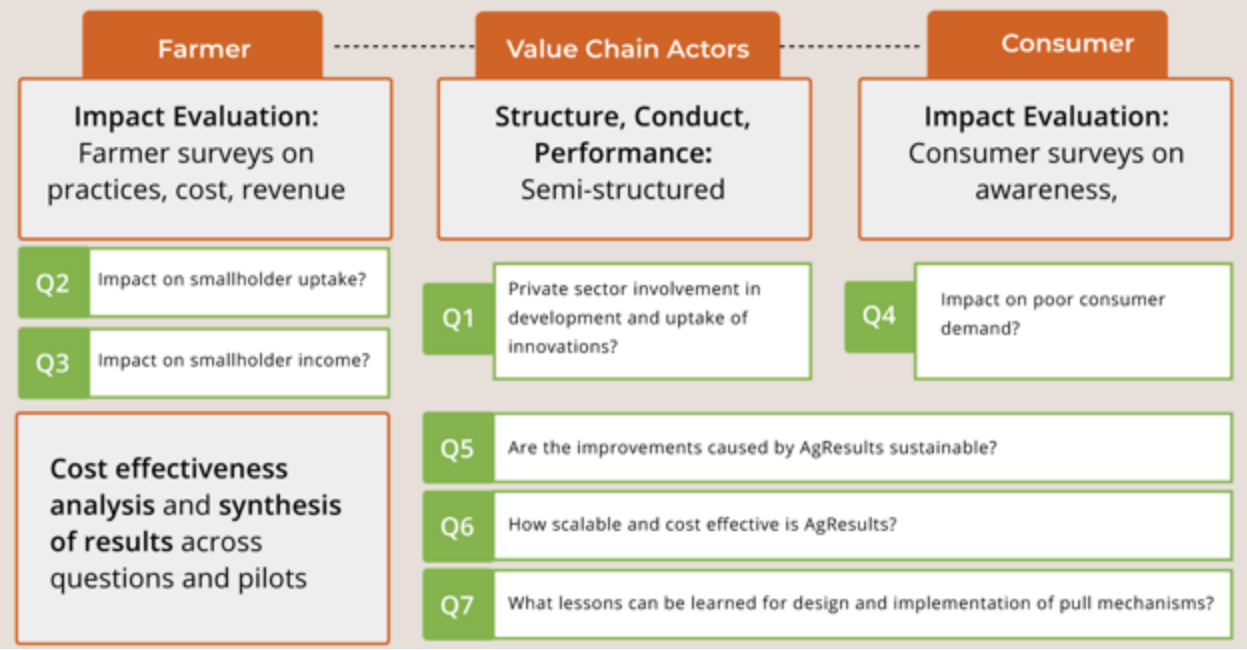
AgResults is a $152 million initiative that uses Pay-for-Results (PfR) prize competitions to incentivize the private sector to invest in high-impact agricultural innovations that help achieve their goals (AgResults, n.d.). One of their key objectives is to test the effectiveness and efficiency of pull financing (i.e., providing incentives/prizes for innovation and/or solving a specific problem) to encourage uptake of innovative agricultural technologies. By linking payments to demonstrated results, it aims to achieve guaranteed impacts and maximum value for money (Davies, 2012).

For program evaluations, the organization also hires external evaluators who peruse a common evaluation framework across all AgResults prize competitions, which allows for comparison

across AgResults' prize competitions and generalization of results and lessons learned. Their approach involves a mix of quantitative and qualitative methods, cost-effective analysis of Challenge Projects and sustainability of results, and comparative analysis of outcomes of other PfR programs (AgResults, n.d.).

The diagram below (Figure 1) outlines their sample evaluation strategy:

Figure 1: Sample evaluation strategy of one of AgResults' programs (AgResults n.d., b)



While appreciating their PfR programs to be an "effective tool," an independent review of four PfR projects by AgResults published in March 2022 suggested some changes to better achieve their programs' desired outcomes. These include structuring the prize to encourage the creation of competitive agricultural markets (especially in developing countries where market development can be very challenging); aligning the prize structure with the development goal of improving smallholder farmers' food security; creating an enabling environment, and underlying assumptions about competitors' response to the prize (Narayan, Geyer, Mainville, & Ness-Edelstein, 2022). While these recommendations are useful, often the program outcomes as opposed to outputs, especially in under- and un-developed markets, are subject to several extraneous factors (such as an enabling environment, institutional barriers, market efficiency etc.).

The emphasis on results-based approach to evaluation for agriculture sector programs is evident in some Government of Canada programs. For example, in 2019, Agriculture and Agri-Food Canada (AAFC) conducted an evaluation of the Canadian Agricultural Adaptation Program (CAAP) (2014-2019) to assess the relevance, performance, design, and delivery of the program during 2014-2019. Their methodology included a document and data review; a literature review;

interviews with AAFC staff, recipients, applicants, project stakeholders, and external stakeholders and experts, and case studies. AAFC developed a logic model to describe the key activities and outputs of the CAAP and the sequence of short-term, medium-term, and long-term outcomes (see Appendix B for details) The evaluation assessed the relevance, performance, design, and delivery of the Program, including lessons learned and recommendations (Agriculture and Agri-Food Canada, 2019).

Similar evaluation approaches have also been utilized for programs under the Agricultural Marketing Act from 2014 to 2019 such as the Advance Payments Program to evaluate their design, delivery, performance, and areas of further improvements (Agriculture and Agri-Food Canada, 2021). This research aims to draw on the knowledge from these recent and relevant public sector (e.g., the Government of Canada agri-sector programs) and private-sector initiatives (e.g., AgResults) in developing the program evaluation framework for IAF.

Summary: Literature Review

In the following table, the key themes from the literature review are aligned with the research questions explored in the following chapters:

Figure 2: Literature Review Themes

Literature Review: Main Themes	Research Questions
Defining evaluation - This thesis uses Patton's definition of evaluation as "the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding" (Patton, 2008, p. 39)	What is program evaluation?
Different types of evaluation - The three main evaluation types are formative (also called process), summative (also called outcome) and impact evaluations.	What are the different evaluation types?
Key elements - Some of key elements of evaluation discussed in this section are program monitoring, accountability, performance measurement and management	What are the key elements of an evaluation?
Evaluation challenges - Two main challenges faced by nonprofits include i) complexity of performance measurement: nonprofits often pursue outcomes that are often difficult to achieve, let alone measure; ii) lack of	What are the program evaluation challenges faced by nonprofits?

targeted research on effective evaluation frameworks for nonprofits compared to the private sector	
Evaluation challenges – agri-sector - Most evaluation programs are accountability-focused and extension-services oriented. Very few of them measure program impacts.	What are some program evaluation challenges faced by nonprofits in the agriculture and/or agri-tech, agri-food sectors?
Overcoming barriers - To address some of these challenges, three core strategies have been highlighted for effective performance measurement for nonprofits that involve (a) nurturing local stakeholder involvement, (b) creating specific goals logically linked to metrics that measure, and (c) continually fine-tuning measures and goals strategically linked to balancing the needs of funding partners.	What are some of the strategies to cope with challenges faced by nonprofits in program evaluation?

Theoretical Framework

Fynn et al. (2020) note that evaluation frameworks facilitate a systematic approach to evaluation. They enable multiple stakeholders to gain a shared understanding of the program and evaluation process, and to identify and agree upon evaluation objectives and methods (Fynn, Wendy, Milton, & Jones, 2020). A theory-based evaluation framework uses a theory of change to draw conclusions about whether and how an intervention contributed to observed results during program evaluation (Treasury Board of Canada Secretariat, 2021). As Kuada (2012) states "Theory provides the language, the concepts, and assumptions that help researchers to make sense of the phenomenon that they seek to investigate." (p. 64). The interest in theory-based approaches has stemmed from the need to measure the impact of public social programs effectively in light of constrained resources (Dauphinee, 2015).

Huey-Tsyh Chen and Peter Rossi developed the concept of theory-based evaluation as the logical extension of quantitative models about identification of variables contributing to the outcomes of a program (Mertens & Wilson, 2019, p. 60). This type of evaluation combines both the practical and scientific aspects of research to formulate evaluation questions that look at the ability of inputs to yield anticipated outcomes (Peer, 2017).

While there is no fixed definition, program theory consists of a set of statements that describe a particular program, explain why, how, and under what conditions the program effects occur, predict the outcomes of the program, and specify the requirements necessary to bring about the desired program effects (Sharpe, 2011). Program theory uses three components to describe the program: the program activities or inputs, the intended outcomes or outputs, and the mechanisms through which the intended outcomes are achieved (Sharpe, 2011). The relationships between these components are often demonstrated by logic models, a visual illustration of a program's

resources, activities, outputs and expected outcomes (Public Health Ontario, 2016). A logic model is a tool to show the intended causal linkages in a program and identify program activities, their intended outputs, and the outcomes (often short, medium, and longer term) (McDavid & Hawthorn, 2010).

There are two main approaches for formulating program theory: stakeholder-based and social science approaches (Rowan, Hogg, Kristjansson, & Dahrouge, 2009). These approaches are very useful for disciplines like the agrisector that often lack sizable prior research because evaluators can still tap into other sources such as program staff, program documents, stakeholder engagement, etc. to develop specific evaluation frameworks tailored to the organizational needs (Donaldson, 2021). The stakeholder-based approach relies on methods such as interviews, focus groups, site visits, and observation (Rowan, Hogg, Kristjansson, & Dahrouge, 2009).

Leeuw (2003) cites three circumstances under which theory-based evaluation is worthwhile. The first is when a program is expected to have a large impact, such as a program that is delivered provincially or nationally. The second is when there are large risks or uncertainty about a program's impact and unintended side effects, for example, new and innovative programs to prevent substance abuse among youth. The third is when there is a strong focus on accountability by the funding partners and/or agencies (p. 17). Most of the Investment Agriculture Foundation's (IAF) programs likely meet these criteria. First, IAF delivers at least 13 province-wide programs related to agri-innovation, agriculture-marketing and climate-change adaptation and mitigation, among others. Secondly, there is an inherent uncertainty in the agriculture sector as weather, commodity prices, government policies, global markets, and other factors that impact farming can cause wide swings in agriculture outcomes and productivity (USDA, 2022). This risk is likely increased when delivering agri-innovation programs that introduce new farm technologies to the farmers (either newly invented or not widely available in the province), which IAF also does. Third, IAF management has indicated an increased demand for accountability from their funding partners that focuses on demonstrating increased value-for-money for their projects, which is also a primary motivation behind conducting this research.

Given these advantages, this thesis will largely draw on the program theory framework because it helps better articulate the "why" and the "what" of the program, i.e., its raison d'être *and* the process to achieve its objectives (Grosvener Public Sector Advisory, n.d.). Moreover, program theory framework is versatile in that it can be useful for a single evaluation, for planning cluster evaluations of different projects funded under a single program, or to bring together evidence from multiple evaluations and research (Rogers, n.d.).

Chapter 3: Methodology and Methods

This chapter identifies the methodological approach, the methods used to gather information, and the ways in which the data was analyzed in this thesis.

The research required and received approval from University of Victoria's Human Research Ethics Board, having the Ethics Protocol Number: 21-0484 (See Appendix A: Certificate of Approval).

Methodology

The research in this thesis is qualitative and uses a program theory framework.

Using program theory, this research develops a conceptual framework around the the main research question: What is the most effective program evaluation framework(s) for IAF?

Jurisdictional Scan

A jurisdictional scan of comparable countries such as New Zealand and the Netherlands is included to document the smart practices prevalent in these countries about evaluation metrics, frameworks, and reporting criteria in the agri-tech and agri-food sectors. Where there is insufficient literature on program evaluation and agri-food and agri-tech industries, the research has explored program evaluation frameworks in similar industries (such as forestry, natural resources, and manufacturing (particularly, food processing) industries), and sectors such as the public and social sectors.

The jurisdictional scan, another approach in this thesis, focuses on identifying, comparing, and summarizing data sources (such as program evaluation frameworks and documents published by government agencies), about effective evaluation criteria, smart practices and evaluation frameworks being used in targeted industries (such as agriculture, agri-tech, and agri-food industries) in the Netherlands and New Zealand. The Netherlands and New Zealand (NZ) both have a large agricultural economy. The Dutch agricultural sector exports some CAD 85 billion worth of agricultural produce annually whereas New Zealand's food and fiber sector's annual exports were forecast at CAD 41.8 billion by June 30, 2022.

As such, IAF identified that these countries could provide some instructive insights from the evaluation frameworks and/or smart practices in the agriculture, agri-food, and agri-tech industries to help identify robust program evaluation criteria and develop the new program evaluation framework. Some of the research themes for jurisdictional scan include program evaluation policy, evaluation frameworks, evaluation smart practices and examples of program evaluation frameworks (where available) in the Netherlands and New Zealand.

Current State Analysis

The research also analysed IAF's existing project activities and financial reporting templates to get a better understanding of their existing program evaluation techniques, frameworks and to identify any gaps.

Methods

Methods are the ways to collect data and this research relied on two primary methods: key informant interviews with IAF staff and a document review of existing IAF project logic models and reporting templates and in relation to the jurisdictional scan, an analysis of evaluation frameworks and/or smart practices in the agriculture, agri-food, and agri-tech industries.

Key Informant Interviews

Four semi-structured key informant interviews form the bulk of the evaluation framework development criteria. The interviewees were selected by the Program Manager at IAF who is also the informal client for this research. The interviewees include the Chief Executive Officer of IAF, two members of project delivery team who are responsible for the design, delivery, and evaluation of IAF programs, and one member of the finance team who is responsible for collecting financial data from the IAF program participants and reporting it to the IAF funding partners.

The interviewees were requested for an interview by the Principal Applicant (the student) through Human Research Ethics Board (HREB) approved email template (see Appendix A for the email template). After the interview invite is accepted, interviewees were provided HREB approved consent form prior to the interviews (see Appendix B for the consent form template). The Interview questions were finalized with feedback from IAF's project lead for this research. The interviewees were provided the finalized interview questionnaire in advance (See Appendix C for the interview questionnaire) to provide their comments on the questions if they chose to do so. Each interview was scheduled for 45 minutes and lasted between 45 minutes to one hour. The four interviews were scheduled on June 9, 10 and 15, 2022, respectively.

These interviews are particularly relevant the research objectives, as they have direct knowledge related to identifying the challenges faced by IAF and their partners in existing program evaluation frameworks and in brainstorming the new program evaluation framework. Therefore, identifying these participants aligns with purposive sampling methods, that is, the intentional selection of informants (in this case, the IAF staff) based on their particular knowledge of, and/or experience with, the focus of inquiry (Robinson, 2014).

Document Review

The document review analyzes IAF's existing activities and financial reporting templates, and logic models. These documents, along with the interviews with IAF staff, help understand the current state of IAF's evaluation practices and identify any gaps in the existing evaluation criteria and practices. In this way, the document review complements interview and jurisdictional scan findings to help suggest ways to build a more effective program evaluation framework.

A jurisdictional scan consisting of document review of publicly available documents in English from the Netherlands and New Zealand is also conducted to find instructive insights from the evaluation frameworks and/or smart practices in the agriculture, agri-food, and agri-tech industries to help identify robust program evaluation criteria and develop the new program evaluation framework. Some of the key search phrases are: "Evaluation policy in the Netherlands" "Evaluation policy in New Zealand", "Agriculture Evaluation Policy in New Zealand" "Agriculture evaluation policy in the Netherlands", "Policy Evaluation Best Practices",, "Agriculture Sector Evaluation Frameworks in the Netherlands", "Program Evaluation Frameworks in the New Zealand", "Government Evaluation frameworks in the Netherlands: best practices", "Dutch Audit Office evaluation policy", Dutch Ministry of Finance evaluation framework", "NZ Treasury Evaluation frameworks", "Social sector evaluation policy in the Netherlands," "Social Sector evaluation in New Zealand," etc.

Data Analysis

The interviews were transcribed and sequenced by thematic areas. A thematic analysis is a qualitative research strategy for identifying, analyzing, and reporting identifiable patterns or themes within data (APA, n.d.). It is a useful method for examining the perspectives of research participants, highlighting similarities and differences, and summarizing large data sets and/or information (Nowell, Norris, White & Moules, 2017). The thematic analysis of interviews along with the secondary research (jurisdictional scan & literature review) helped to identify robust program evaluation criteria and develop the new program evaluation framework.

Limitations and Strengths

Limitations

While primary research (interviews with IAF staff) informed the key aspects of developing the program evaluation framework, the research findings were limited by the quality and quantity of secondary sources analyzed during the jurisdictional scan. As discussed in the jurisdictional scan chapter, there seems to be a dearth of existing program evaluation frameworks specific to the agriculture or agri-tech or agri-food sector, in general**.** In their absence, the search expanded to include program evaluation frameworks in comparable sectors such as the public and social sectors (where available). Another limitation encountered during the jurisdictional scan of the

Netherlands was that the evaluation frameworks being used there were either publicly unavailable or unavailable in English, with the exception of the Netherlands Ministry of Foreign Affairs' "Evaluation policy and guidelines for evaluations" (2009), which is included. To compensate, the research expanded to include some recent and relevant scholarly articles on program evaluation in the Netherlands. Lastly, the research is aimed to contribute to the broader literature on effective evaluation practices and frameworks for programs and organizations (such as IAF) that fund and support agriculture, agri-food, and agri-tech initiatives, and if possible, how these practices and frameworks can be adapted for different geographic and organizational contexts.

Strengths

One of the key strengths of the thesis is that it will draw on a multi-faceted methodological approach involving both primary sources such as interviews with IAF staff and secondary sources such as jurisdictional scan of program evaluation frameworks in comparable countries, and the document review of IAF's existing evaluation practices. This approach coupled with the literature review, helped tap into both global evaluation practices, models and frameworks, and IAF staff's insights to help compile smart practices and effective evaluation criteria that can be perused by IAF in their program evaluations.

Chapter 4: Current State Analysis - Interview Results

This chapter describes the key findings from the four semi-structured interviews conducted with IAF staff. The interviews have been transcribed and the results are described in a question-answer format, as generalized responses, without attributing individual responses to their respective respondents.

Theme 1 - Background Information

Participants were asked the following question: **Could you please briefly describe your position and role at IAF?**

The interviewees include the Chief Executive Officer of IAF, two Program Managers who are responsible for the design, delivery, and evaluation of IAF programs, and one Policy Analyst from the Finance Team who is responsible for collecting financial data from the IAF program participants and reporting it to the IAF funding partners.

Theme 2 - Data Collection and Developing the Baseline

Participants were asked the following question: **What data is IAF gathering to assess programs and is it the right data?**

Most programs have more applicants than the funding available so for most of IAF's programs, they do a competitive scoring on a bunch of indicators and assign a numerical score between one and five for each of those components. Some of it could be based on how well the project aligns with the program goals. For instance, if it is an innovation program, IAF evaluates if it was innovative; what is novel about what the client is proposing, etc.

Participants were asked the following question: **What additional data and information should be collected to improve IAF's program evaluation, and why would these data/information improve program evaluation?**

Some improvements that were noted by the participants include:

- **Need for more qualitative data**: IAF is good at capturing quantitative metrics, but they need to integrate those softer narrative pieces to create more fulsome overview of the programs.
- **Less reliance on client-supplied data**: clients are required to submit activities and financial reporting templates for most projects, so, IAF often relies on client supplied information to determine if the project achieved its intended targets. The proof could be invoices for purchases, YouTube videos of project sites or new machinery, etc. There is a need for more field visits for some projects, especially on-farm projects.

Participants were asked the following question: **What is IAF's current capacity for collecting and analyzing the aforementioned additional data? What is needed to improve IAF's capacity for collecting and analyzing this data?**

The key responses for this question include:

- **Capitalize on data visualization software:** Internal dashboards such as Power BI or Tableau (interactive data visualization software with a primary focus on business intelligence) can provide real-time reporting on the program budget (e.g., proposed versus actual budget) and help capture project activities to better monitor if a project is on track and to analyze how clients are engaging with the programs throughout its duration, without having to wait till project completion. It is often a short window when IAF is asked to deploy a program so it might be difficult to build these dashboards. Nonetheless, if IAF was developing its own programs, the software could be very useful in terms of communicating within, and also to quickly tell a story to the funding partners or provide information to their public partners.
- **Need dedicated funding for evaluations**: Most programs do not have an administration fee to conduct evaluations and therefore IAF conducts evaluations based on their resource capacity. Secondly, not all programs ask for an evaluation but rather ask for specific reporting metrics that they provide them. It could be a one-page document summarizing the highlights of the program.
- **Need for more integration between the Finance and Project Delivery teams**: There is sometimes separation between the finance and project delivery team and they need to be more integrated. Towards the end of the project when IAF analyzes project documents, some documents submitted by clients can be both relevant to the Finance team and the Project Delivery team so there is potential for integration between the two for improved program delivery and evaluation.

Theme 3 - Current State – Performance Indicators

Participants were asked the following questions: **What are some of the key existing input and output indicators of success programs delivered by IAF?**

Most of the sales-related indicators are summative in approach; however, for most programs, the indicators are set by the funders who are "quite prescriptive" about the specific metrics they want measured such as the following:

- how much money needs to be spent on X number of projects per intake or per year?
- how much was spent?

Participants noted that program success is often measured on how much of the approved funds were spent on project activities.

Participants were also asked: **Which indicators are common among the programs, and which are program-specific? Are there any supporting data/info/docs that can help me better understand the process?**

Program indicators are mostly program-specific and set by the funding partners, except for sales-related programs, which have common indicators. For example, one interview highlighted that a common indicator could be: did IAF spend all the project money on the approved activities?

For some programs in which IAF plays a developmental role, IAF identifies industry needs and gathers feedback from relevant stakeholders to develop program logic models for those programs. One participant also noted IAF has also developed activities and financial reporting templates.

Participants were asked the following question: **What are key data/information sources for the indicators?**

The interview participants stated that the key data sources are the financial reporting templates, which document how the project money is spent, and the project activities reporting templates, which describe the activities that were undertaken with the project funding (e.g., attending trade shows, buying farm machinery, etc.) One interviewee highlighted that the activity reporting is still quantitative in nature versus qualitative but it is not thematically coded to observe trends over time. Other data sources noted by the participants include invoices for purchases made as part of the programs.

Two participants highlighted that data collected is often labor intensive and requires considerable staff time to sort through and report to the funding partners. It was also reported that sometimes IAF makes short reports instead of detailed program evaluations) about project highlights (depending on the project) for funders but it is not part of their typical reporting structure.

Participants were asked the following question: **How were the indicator systems developed? Did IAF draw upon other indicator systems and frameworks?**

The reporting templates were developed by the finance team to document spending of the clients. The activities reporting template was developed by the project delivery team. There are also online end-of-program surveys sent to the clients which are mandatory to receive funding. These templates ask questions such as did the project obtain footholds? Did it increase jobs? How did the project support economic recovery of BC’s agricultural landscape? etc.

Theme 4 - Logic Models and Evaluation Frameworks

Participants were asked the following question: **To what extent are the inputs, activities, outputs, and outcomes in the evaluation framework(s) consistent with IAF program objectives?**

IAF has its own internal organizational goals about how they are supposed to function but the program indicators and objectives are set by the funding partners for the most part and IAF's role is to help implement those objectives. Some examples of questions about program objectives are: how is the program initiative expected to lead to certain outcomes or impacts, did IAF do a good job of identifying the underlying assumptions of the programs, etc.? Program objectives are not absolute and can be affected by extraneous factors such as COVID, which diminished the targets of some marketing and sales programs.

Participants were asked the following question: **Are you aware if logic models or other tools were used in drafting the existing evaluation framework?**

The interview participants noted that for some programs where IAF plays a developmental role, they develop logic models to outline inputs, activities, and outputs. Having said that, outcomes (often medium- and long-term) are difficult to calculate as they require data and follow-up several years after project completion which is difficult to achieve.

Theme 5 - Evaluation Success and Client Satisfaction

Participants were asked the following question: **Are participants being reached as intended? How are program clients contacted regarding the success of their projects? Is this an effective method of communication?**

Participants (recipients of IAF programs) are approached by email or phone and provided reporting templates for financial and activities reporting, which is usually a Word or an Excel document. For some programs, IAF conducts baseline and end-of-the-program surveys through online forms (such as Microsoft forms), which allows building out the questions based on the reporting requirements of their funding partners and what they're looking to measure at a project level. Those surveys are sent via email to the participants.

Participants were asked the following question: **Do you measure client satisfaction? If so, how? (i.e., what instruments do you use?)**

Client satisfaction is measured through end of program surveys. These are very generic surveys that ask about the application processes through the questions: how was the experience of applying to the program, how was the program delivered, how they found communication on the website, were they able to access materials or information on the website, etc.? The feedback is evaluated by the Program Delivery team. The end of program surveys is not mandatory and therefore their response rate is often low (around 25%).

Participants were asked the following question: **What areas (if any) have the program clients indicated dissatisfaction or desire for improvement?**

The IAF used to have paper-based surveys and application documents but they have now moved most of their files to email, to make it less cumbersome. They are now developing online forms and survey tools (to completely do away with email-based forms) to make the process more efficient. There is an open field at the end of the survey where the program participants can indicate dissatisfaction or areas for improvement. Having said that, most clients that face any issue already communicate those concerns through email or phone well before the surveys.

Theme 6 - Strengths and Challenges – IAF Program Evaluation

Participants were asked the following question: **What are the key existing challenges/gaps in IAF's ability to evaluate the programs it delivers?**

One of the main challenges identified was capacity constraints at IAF, i.e., lack of dedicated staff and fiscal resources for program evaluation. Another challenge is to improve data systems at IAF but that is just inherently iterative as technology keeps getting better. It is, however, more about the balancing it by recognizing the clients' capacity, as some clients (mostly small farmers) may not be tech-savvy or not have strong internet access to cope with advanced technology. A third challenge is that some programs do not have a clear logic model or well-developed program objectives so that makes it hard to measure the impacts of the program.

It was also highlighted that there was lack of focus on storytelling based on the data collected and that needs to be built in the program evaluation. Whereas data collection and lack of baseline data is also an issue. For example, some programs have baseline data but others do not which makes it difficult to establish cause and effect relationships. For example, if a marketing program is intended to increase sales, having sales figures before the start of the program would be helpful in determining how much (if at all) the sales increased after the marketing program was delivered.

The interviewees also highlighted fewer field visits as a problem, as IAF mostly relies on documentary evidence. IAF does this for the Buy BC program however, not all projects require field visits. For trade shows, IAF has templates to document the learnings.

Moreover, clients' organizational capacity was also highlighted as an issue as most times, it is only one person (i.e., the farmer) handling the entire project on their side.

Finally, interviewees highlighted having lesser reporting time as a challenge. IAF project funding is tied to the fiscal year and the impacts of some projects take longer to measure than the funding or reporting cycle. For example, in the Bee BC Program, effects of developing a new overwintering technology for bees needs to be tested once it's warm enough for the bees to come

out of the hives and get active. The project reporting, however, needs to be completed by February 15 because it is the end of the fiscal year. In northern BC, where it is still cold, the bees have not woken up by then so, it is difficult to know the full impacts of a project at the time of reporting. The clients often also move on after the project so, it is difficult to approach them a year or two after project completion.

Theme 7 – IAF and Program Evaluation

Participants were asked the following question: **What is IAF hoping to achieve through more effective program evaluation practices?**

The main themes identified in the responses are:

- **Need for formative evaluations:** There is a need to integrate the formative evaluation component with the summative one (that already exists), to evaluate how the program is performing, and if it is reaching its intended outcomes.

- **Establishing a feedback loop and making reporting formats client-friendly:** Some interviewees called for establishing a feedback loop. For instance, if a client is facing an issue with the application process or an administration component, they can report it in an efficient way. The reporting process for that needs to be easier. If the reporting format is difficult to navigate and fill, it becomes more difficult to extract proper information from the client, analyze it in real-time and make those changes.

- **Need for more targeted and goal-based programs**: Most programs have very broad objectives and they are also very activity-based and therefore they receive a lot of applications, most of which are ineligible. There is a need for more targeted programs with few but measurable goals that can help attract better applications and generate more quality data for impact measurement. Last year, IAF had to turn down around 90% of applicants after selecting only the top ones, which is a waste for people who applied and got rejected. To avoid that, IAF needs programs with targeted objectives.

Participants were also asked the following question: **What elements would make for a comprehensive and replicable evaluation framework at IAF?**

Some improvements that were highlighted are:

- **Need for baseline data and field-visits:** Having baseline data can help establish cause and effect relationships and field visits can help better monitor program impacts.
- **Program-specific objectives:** Each program should have a well-defined logic model and specific targets to measure program impact, and to discourage ineligible applicants.

Continuously improve "best-practices": IAF has standard operating procedures to provide quality services to the clients but that needs to be continuously updated through feedback from clients.

In closing, participants were asked the following question: **Is there anything else I should be asking?**

Respondents noted that apart from delivering programs funded by its partners, IAF has also, in the past, designed their own programs from repurposed funding left over by some of the other programs. In future, when they plan to do it more, it would be important to find out how to how to go about it. For example, how to be accountable? How to collect the information and how to tie that into the organizational goals? It would be important to not repeat the same mistakes with program design and find ways to measure outcomes. This requires sufficient timeline to capture some impacts and do a review six months, a year, or three years after project completion to measure long term outcomes and to analyze any unanticipated consequences. That would help demonstrate value-for-money.

Summary of Interview Findings

In interviewing IAF staff, they noted that the current key sources for IAF for collecting program data are financial reporting templates, which document how the project money is spent, and the project activities reporting templates, which describe the activities that were undertaken with the project funding. The IAF also sends a voluntary end-of-program survey to measure client satisfaction about working with IAF and that has a 25% response rate and IAF develops logic models to outline program goals, inputs, activities, and outputs for some programs in which IAF plays a developmental role. Calculating outcomes (often medium- and long-term), are difficult as they require data and follow-up several years after project completion which is difficult to achieve, since most IAF projects are tied strictly to their funding or budget cycle.

Some of the key challenges/gaps noted by the participants in IAF's ability to evaluate the programs it delivers are: i) Lack of dedicated staff and fiscal resources required for program evaluation; ii) unavailability of well-developed logic models for some programs; iii) Need for integrating storytelling based on the data, in program evaluation, along with quantitative measures; iv) Insufficient baseline data to compare data before and after a program is delivered; v) Lack of field-visits to monitor and verify progress of some programs (particularly on-farm projects); vi) Insufficient time to measure project impacts, as most projects are tied to the funding cycle and IAF needs to report back to the funders before the program impacts become fully apparent.

IAF staff identified a few measures to address their challenges and develop effective program evaluation practices at IAF. These are: i) Integrating formative evaluation component in their evaluation practices; ii) Establishing a feedback loop and making reporting templates more

client-friendly; iii) Developing more targeted and goal-based programs to receive relevant and eligible applications and generate more quality data for impact measurement; iv) Having baseline data and conducting field-visits to help better monitor program outcomes; and V) Creating program-specific logic models. The interviewees also highlighted that IAF's capacity for collecting and analyzing program evaluation data could be improved by i) Capitalizing on data visualization software such as Power BI or Tableau to provide real time reporting on the program activities, budget, and related data to better monitor if a project is on track; ii) Including a dedicated administration fee for IAF for program evaluations; and iii) Improving integration between the Finance and Project Delivery teams at IAF for improved program delivery and evaluation.

Chapter 6: Current State Analysis - IAF Document Review Findings

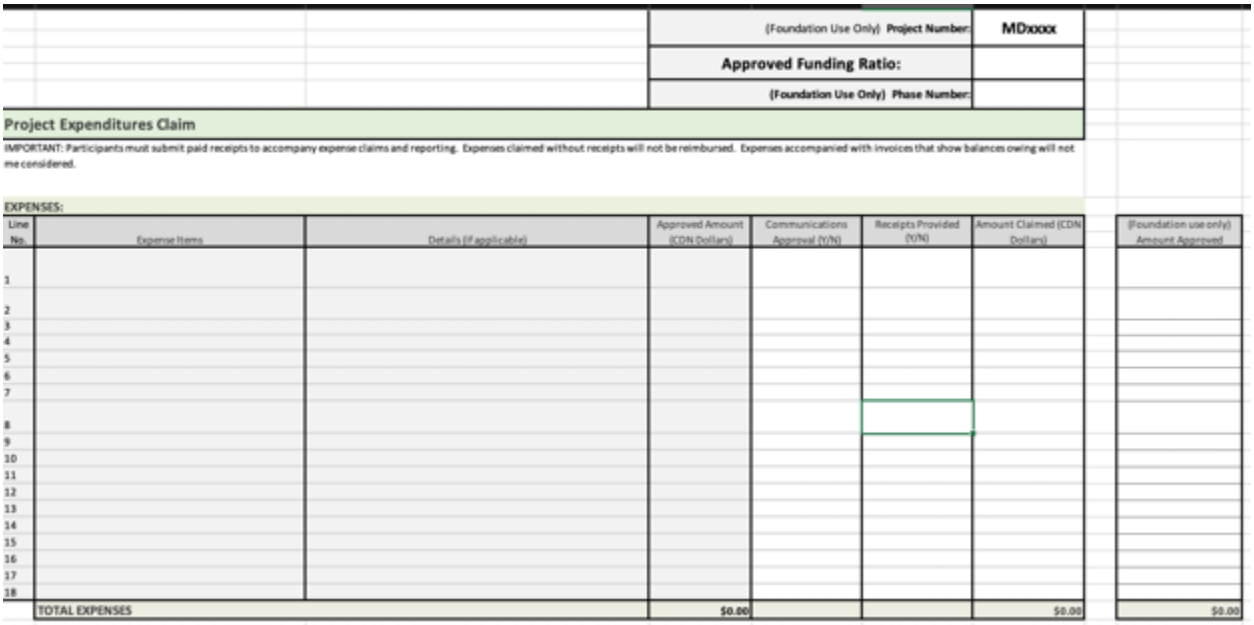
This chapter describes IAF's existing activities and financial reporting templates, and logic models shared by IAF staff from the Project Delivery and the Finance teams. The reporting templates include the "BC Agri-food and Seafood Market Development Program: Financial Reporting Template", "BC Agri food and Seafood Market Development Program: Project Reporting Template and the "MD Multi-activity Reporting Template". The logic models include the "Buy BC Logic Model", "CAP Market Development Logic Model", and the "Farmland Advantage Program Logic Model". The activity and financial reporting templates track the eligible activities part of the IAF funding programs, whereas the financial reporting templates keep a track of funds dispensed to IAF program recipients. The information gathered from these templates is reported back to IAF's funding partners.

These templates, along with the interviews with IAF Staff, therefore, help understand current state of IAF's reporting and evaluation practices and identify any gaps in these practices. In this way, the document review is designed to complement the interview and jurisdictional scan findings to help suggest ways to build a more effective program evaluation framework.

Financial and Project Reporting Templates

Example 1 - BC Agri-food and Seafood Market Development Program: Financial Reporting Template

This is a financial reporting template, a MS Excel spreadsheet, provided to IAF program recipients of the BC Agri-food and Seafood Market Development Program. It consists of two tabs: Instructions and Project Expenses. The Instructions tab outlines that it is mandatory for IAF recipients to submit the completed templates and provides instructions on how to document project expenses in the template. The second tab (Figure: 5) comprises of a line-item view to record all the eligible expenses that are part of the program. It asks program recipients to list the expense items and provide item details (e.g., eligible items could be a machinery purchase) and the amount it cost (in CAD). It also asks for Yes/No to answers to whether the purchase was approved by IAF and if the recipient is providing purchase receipts. Finally, it asks if the claimed/purchase amount is different from the approved amount. This is repeated for all line items which are then summed to calculate total amount spent. As mentioned in the interviews, IAF uses this template to keep the record of project funding and report back to the funders.



				(Foundation Use Only) Project Number:		MDxxxx		
				Approved Funding Ratio:				
				(Foundation Use Only) Phase Number:				

Project Expenditures Claim

IMPORTANT: Participants must submit paid receipts to accompany expense claims and reporting. Expenses claimed without receipts will not be reimbursed. Expenses accompanied with invoices that show balances owing will not me considered.

EXPENSES:

Line No.	Expense Items	Details (if applicable)	Approved Amount (CDN Dollars)	Communications Approval (Y/N)	Receipts Provided (Y/N)	Amount Claimed (CDN Dollars)	(Foundation use only) Amount Approved
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
	TOTAL EXPENSES		$0.00			$0.00	$0.00

Figure: 3, a snapshot of program-specific financial reporting template.

Example 2 - BC Agri-food and Seafood Market Development Program: Project Reporting Template

This is also an MS Excel spreadsheet that is provided to recipients of the BC Agri-food and Seafood Market Development Program. It is a project reporting template that tracks project activities, expenses (similar to the financial reporting template for the same project), and communications approval from IAF for the activities undertaken and purchases made as part of the program. It consists of four tabs: Instructions, Financial Reporting, Activity Reporting, and Communication Approval. The first two tabs are similar in structure to the financial reporting template of the same program (Figure 3). The Activity Reporting tab has three sections.

The first section asks participants to select which areas they completed project activities in? such as Tradeshows and/or Buyer's Meetings, In-Store Promotions/Demonstrations, etc. The section also asks them to specify if all project activities were completed as planned? And if not, provide reasons for not completing them. The next section is on “Project Objectives & Results” and asks questions such as “Did your funded project focus on entering a new market?”, “Did your funded project focus on diversifying an existing market?” For last question, the onus seems to be on the program recipients to demonstrate if the funding helped them diversify an existing market. This is consistent with two of the interview findings that project success is often measured in terms of activities completed rather than measuring the *impact* of those activities, and that it mostly relies on client-supplied information. The full list of questions in this section is shown in Figure 4.

Project Objectives & Results				
Market Research Activities: Information and Intelligence Products				
	Did your funded project focus on entering a new market?			
	Did your funded project focus on diversifying an existing market?			
What type of product(s) were the focus of your Market Research activities?				
Which geographic regions did you target within Canada (City, Province)?				
What geographic regions did you target outside Canada (Region, Country)?				
		Canada	International	Total
Number of Market Research Studies				0
Number of Market Analysis Reports				0
Number of Consumer Analysis Reports				0
Number of Competitive Analysis Reports				0
Number of Trade Statistics Reports (Purchased)				0
Number of Sales Data Reports (Purchased)				0
Other (please specify name & number produced)	*(type 'other report' name here)*			0
Market Readiness Activities: Market Development Planning & Skills Training				
Which geographic regions did you target within Canada (City, Province)?				
What geographic regions did you target outside Canada (Region, Country)?				
What type of product(s) were the focus of your Market Research activities?				
		Canada	International	Total
Number of Marketing Plans/Strategies created				0
Number of Export Plans/Strategies created				0
Number of Branding Plans/Strategies created				0
Other (please specify)				0
Number of eligible Market Skills Training courses completed				
Number of Marketing Staff who completed coursework				

Figure 4: Questions asked in the Program Objectives and Results section of the project reporting template

The last two sections of the Activity Reporting tab ask about the specific export development activities and the sales resulting from the project (Figure 5). It also provides a section for program recipients to provide additional comments.

The Communications Approval tab reads as a checklist of activities and line items that were performed or purchased, respectively, as part of the program and whether IAF approval was received for them. Overall, this reporting template is more comprehensive than the Financial Reporting template of the same program as it combines both financial Reporting and project activities reporting.

Export Development Activities			
What geographic regions did you target **within** Canada?			
What geographic regions did you target **outside** Canada?			
	Canada	International	Total
Number of Tradeshows (In-Person)			0
Number Tradeshows (Virtual)			
Number of Trade Leads Collected			0
Number of buyers meetings, trade missions and/or other events			0
Number of Product Demonstrations (In-Store)			0
Number of Marketing Collateral/Promotional Tools created (quantified by type)			0
Number of Ads (print, digital, broadcast) created			0
	Canada	International	Total
Sales resulting from your Funded Project. This section is not required for Not-for-Profit Organizations or Industry Associations.			
	Canada	International	Total
Number of New Products Introduced			0
Incremental increase in sales ($)			$ -
Incremental increase in sales volume			0
Number of new sales agreements			0
Value of new sales agreements ($)			$ -
Total estimated sales anticipated (within six months of completion) ($)			$ -
Additional Comments			
Please provide any additional comments regarding your funded project. This could include what went well, challenges, lessons learned or project reach and impact.			

Figure 5: Information asked in the Activity Reporting Template

The MD Multi-activity Reporting Template is very similar to the BC Agri-food and Seafood Market Development Program: Financial Reporting Template and is only different in that apart from documenting line-item expenses, it also asks for research, training, and marketing activities' expenses.

Logic Model Templates and Frameworks

Example 3 - Buy BC Logic Model

The Buy BC program logic model (Figure 8) shares three key objectives of the Buy BC program:

- Increase the availability of B.C. foods at grocery stores and markets,
- Motivate consumers to purchase B.C made foods and
- Promote B.C. made foods across the province

The logic model contains five set of activities to achieve these objectives such as government-led advertising, consumer engagement & local food literacy, retail & restaurant promotions, matchmaking events, and industry-led marketing campaigns. These activities are linked to specific outputs which aim to translate into immediate and long-term outcomes (Figure 8). It seems, though, that the list of indicators to measure these outcomes is missing. For instance, one of the immediate outcomes is that Consumers recognize the Buy BC logo as key certification mark to identify B.C. food and beverages. It does not however, mention *how* the program intends to measure consumer recognition. Without this information, it is difficult to measure program impact. For example, would it be done through consumer surveys or focus groups? The same can be said of other outcomes, where the link between outputs and outcomes, often bridged through performance measurement indicators seems to be missing but is essential for outputs to translate into outcomes (Martin, 2019).

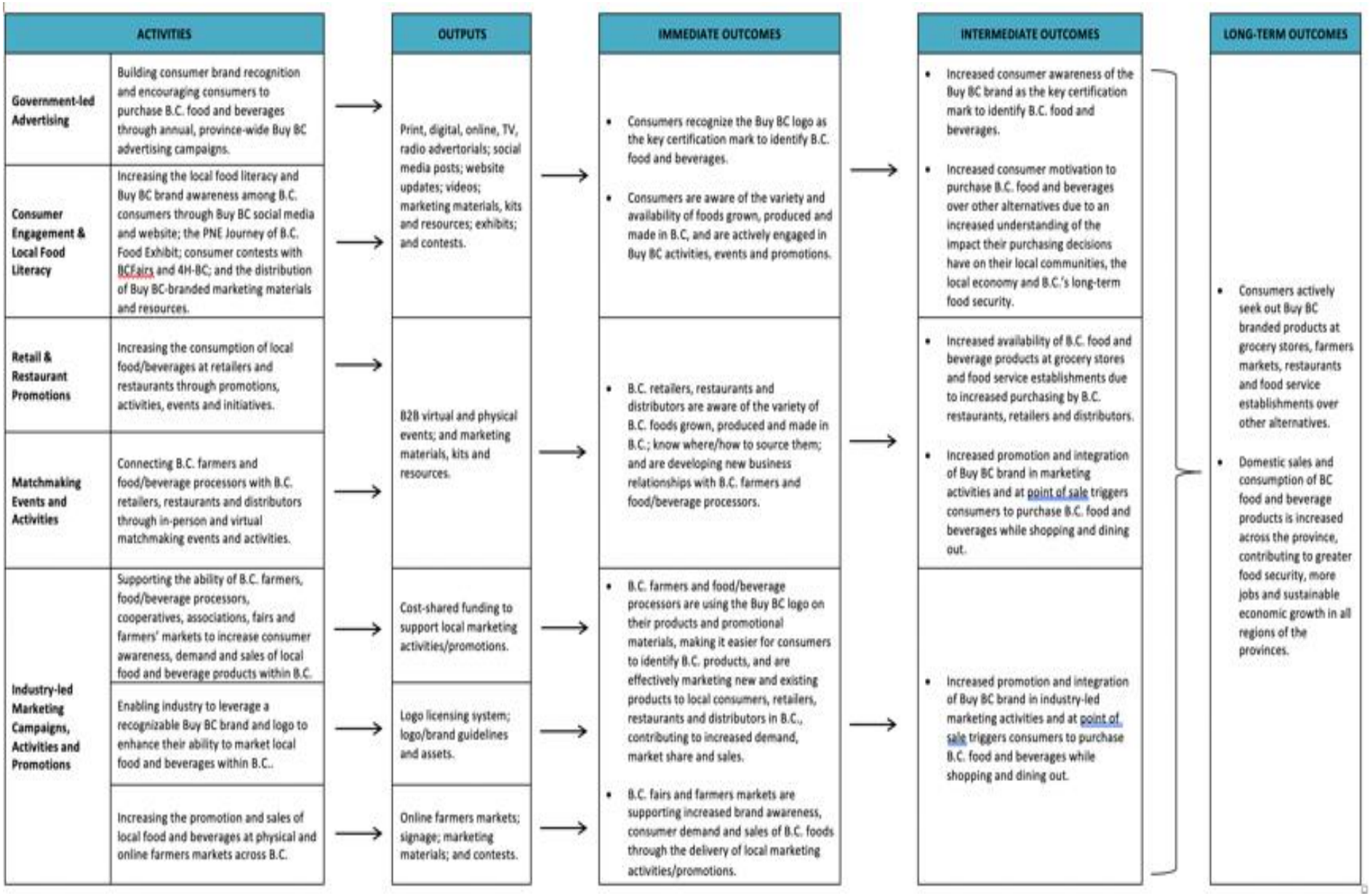

ACTIVITIES
OUTPUTS
IMMEDIATE OUTCOMES
INTERMEDIATE OUTCOMES
LONG-TERM OUTCOMES
Government-led Advertising
Building consumer brand recognition and encouraging consumers to purchase B.C. food and beverages through annual, province-wide Buy BC advertising campaigns.
Consumer Engagement & Local Food Literacy
Increasing the local food literacy and Buy BC brand awareness among B.C. consumers through Buy BC social media and website; the PNE Journey of B.C. Food Exhibit; consumer contests with BCFairs and 4H-BC; and the distribution of Buy BC-branded marketing materials and resources.
Retail & Restaurant Promotions
Increasing the consumption of local food/beverages at retailers and restaurants through promotions, activities, events and initiatives.
Matchmaking Events and Activities
Connecting B.C. farmers and food/beverage processors with B.C. retailers, restaurants and distributors through in-person and virtual matchmaking events and activities.
Industry-led Marketing Campaigns, Activities and Promotions
Supporting the ability of B.C. farmers, food/beverage processors, cooperatives, associations, fairs and farmers' markets to increase consumer awareness, demand and sales of local food and beverage products within B.C.
Enabling industry to leverage a recognizable Buy BC brand and logo to enhance their ability to market local food and beverages within B.C..
Increasing the promotion and sales of local food and beverages at physical and online farmers markets across B.C.
Print, digital, online, TV, radio advertorials; social media posts; website updates; videos; marketing materials, kits and resources; exhibits; and contests.
B2B virtual and physical events; and marketing materials, kits and resources.
Cost-shared funding to support local marketing activities/promotions.
Logo licensing system; logo/brand guidelines and assets.
Online farmers markets; signage; marketing materials; and contests.
Consumers recognize the Buy BC logo as the key certification mark to identify B.C. food and beverages.
Consumers are aware of the variety and availability of foods grown, produced and made in B.C, and are actively engaged in Buy BC activities, events and promotions.
B.C. retailers, restaurants and distributors are aware of the variety of B.C. foods grown, produced and made in B.C.; know where/how to source them; and are developing new business relationships with B.C. farmers and food/beverage processors.
B.C. farmers and food/beverage processors are using the Buy BC logo on their products and promotional materials, making it easier for consumers to identify B.C. products, and are effectively marketing new and existing products to local consumers, retailers, restaurants and distributors in B.C., contributing to increased demand, market share and sales.
B.C. fairs and farmers markets are supporting increased brand awareness, consumer demand and sales of B.C. foods through the delivery of local marketing activities/promotions.
Increased consumer awareness of the Buy BC brand as the key certification mark to identify B.C. food and beverages.
Increased consumer motivation to purchase B.C. food and beverages over other alternatives due to an increased understanding of the impact their purchasing decisions have on their local communities, the local economy and B.C.'s long-term food security.
Increased availability of B.C. food and beverage products at grocery stores and food service establishments due to increased purchasing by B.C. restaurants, retailers and distributors.
Increased promotion and integration of Buy BC brand in marketing activities and at point of sale triggers consumers to purchase B.C. food and beverages while shopping and dining out.
Increased promotion and integration of Buy BC brand in industry-led marketing activities and at point of sale triggers consumers to purchase B.C. food and beverages while shopping and dining out.
Consumers actively seek out Buy BC branded products at grocery stores, farmers markets, restaurants and food service establishments over other alternatives.
Domestic sales and consumption of BC food and beverage products is increased across the province, contributing to greater food security, more jobs and sustainable economic growth in all regions of the provinces.

Figure 6: Buy BC Program Logic Model

The CAP Market Development Program Logic Model (Figure 6) and the Farmland Advantage logic model (Figure 7) both follow a similar format and seem to be missing the critical link: performance indicators (e.g., surveys, interviews, etc.) to measure and ascertain if the program activities and outputs correspond to their intended outcomes. Performance indicators also help underline any issues in program design and delivery to help refine it for future runs.

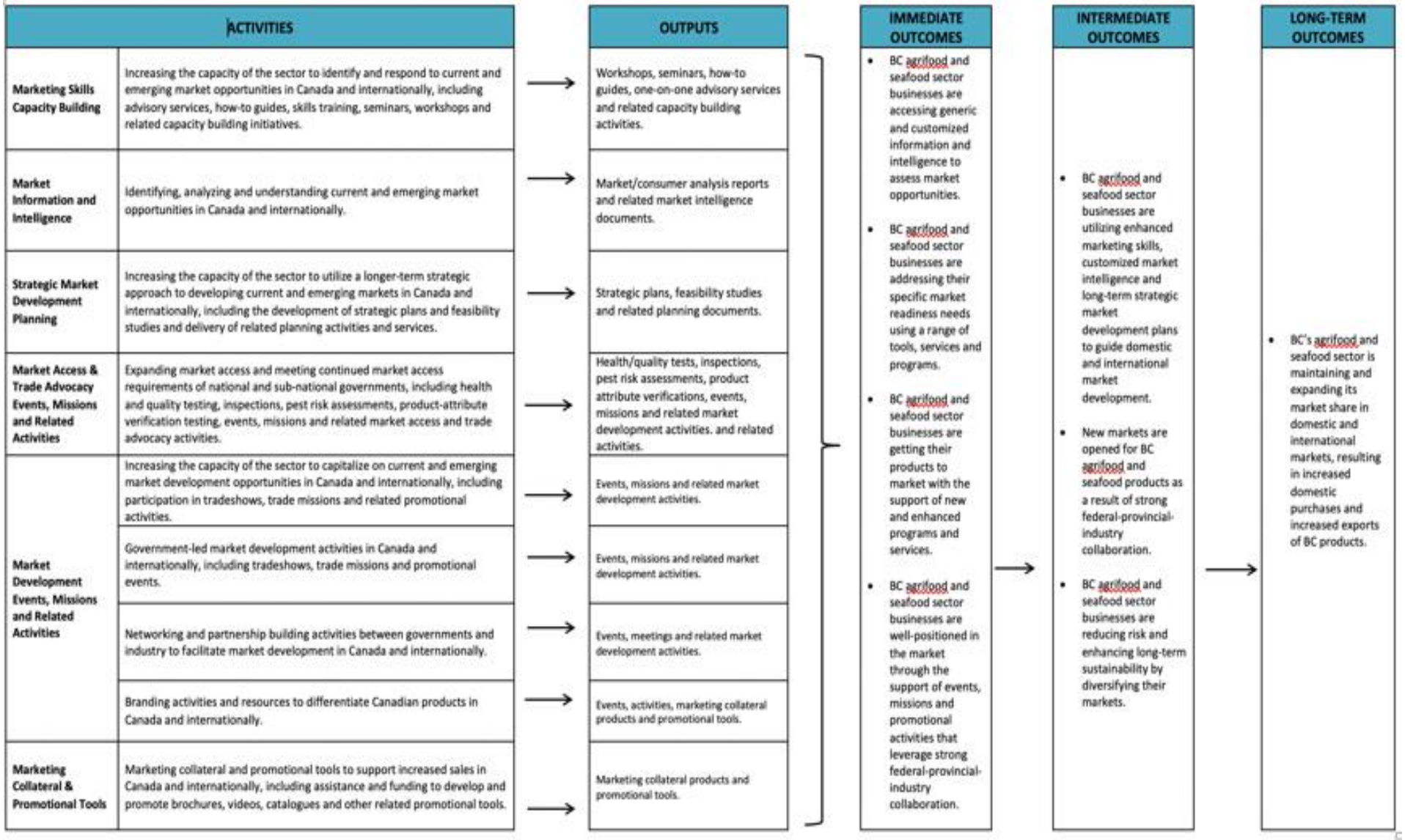


Figure 7: CAP Market Development Program Logic Model

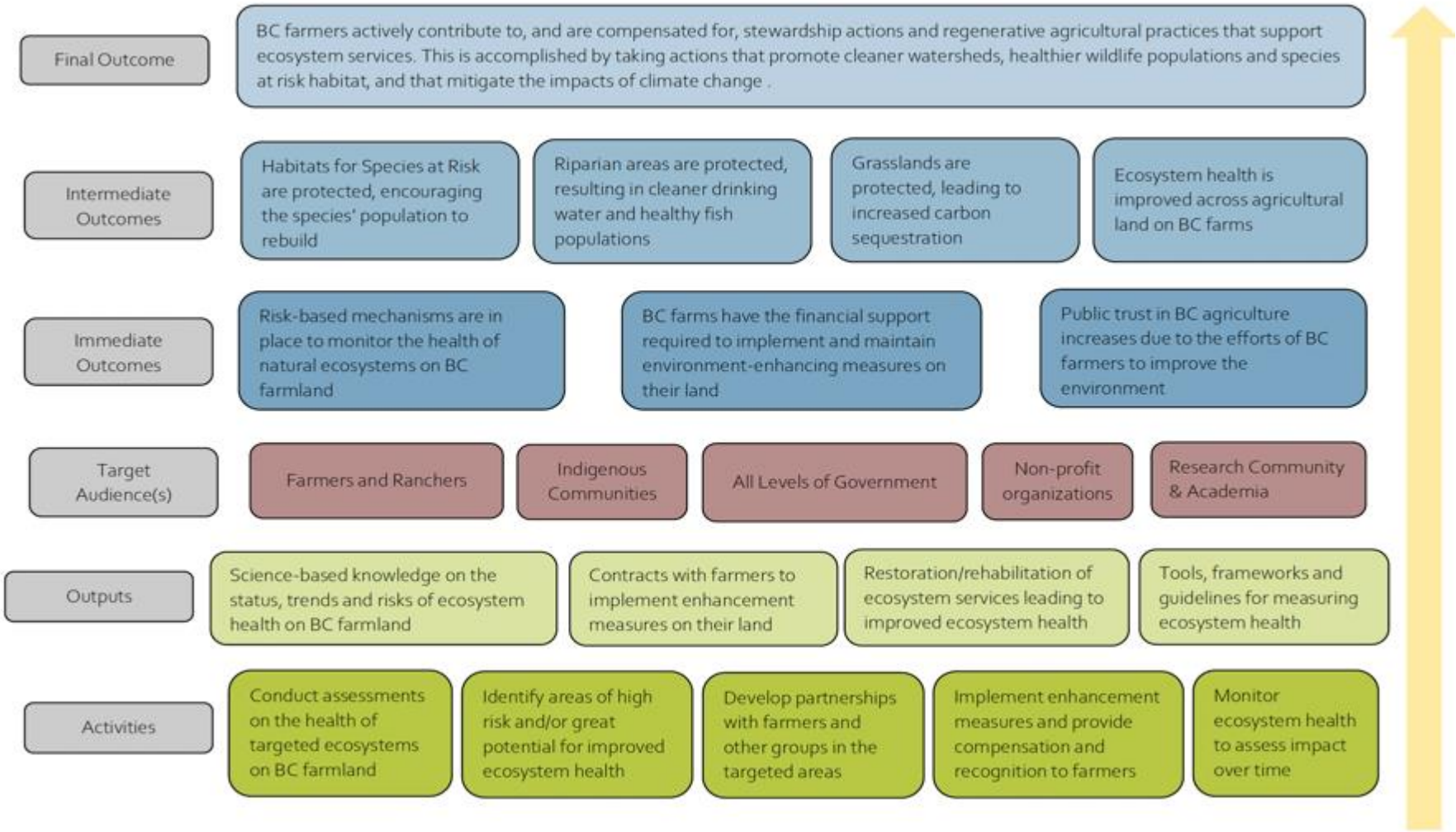


Figure 8: Farmland Advantage Program Logic Model

Summary of the IAF Document Review chapter

This chapter summarizes IAF financial and project reporting templates and logic models to help understand current state of IAF's reporting and evaluation practices and identify any gaps in these practices. The project and financial reporting templates provide detailed overview of eligible activities undertaken by IAF program recipients and the funds dispensed by IAF.

One of the key findings is that the onus in reporting templates seems to be on the program recipients to demonstrate if the funding helped them achieve the intended objectives. This is consistent with the interview findings that project success is often measured in terms of activities completed rather than measuring the *impact* of those activities, and that it mostly relies on client-supplied information.

Another key finding is that for logic models, it is difficult to measure program impacts without performance indicators (e.g., surveys, interviews, etc.) to measure and ascertain if the program activities and outputs correspond to their intended outcomes. Performance indicators also help underline any issues in program design and delivery to help refine it for future runs.

Chapter 5: Jurisdictional Scan Findings

The jurisdictional scan focuses on identifying, comparing, and summarizing data about effective evaluation criteria, smart practices and evaluation frameworks used in targeted industries (such as agriculture, agri-tech, and agri-food industries) in the Netherlands and New Zealand.

The two jurisdictions were identified by IAF based on similarities in industry, demography, and overall structures and processes. For example, both of these countries are not only large agrarian economies, New Zealand also shares certain topographic and demographic similarities with British Columbia that may make adoption and/or replication of certain program evaluation knowledge and frameworks more conducive for B.C. Similarly, the Netherlands is renowned for being a leader in agri-innovation and farm technology space for decades so, it can provide the most recent and relevant expertise for IAF. Moreover, the focus on these two jurisdictions was also inspired by a need for more global perspectives (wherever possible), beyond those already existing in other Canadian jurisdictions such as Ontario.

The purpose of this chapter is to identify robust program evaluation criteria and practices in use in these countries that can help develop the new program evaluation framework for IAF.

The Netherlands:

The Netherlands is a global leader in agri-food innovation owing to its world-class research institutes, agri-innovation practices and public-private partnerships between science, industry, and government – from farm to fork. The Dutch agricultural sector exports some CAD 85 billion worth of agricultural produce annually (17.5% of total Dutch exports), second only to the U.S (Government of the Netherlands, n.d.). About 15 of the top 20 biggest agri-food companies have major production or R&D sites in the Netherlands, including Nestlé, AB InBev, Coca-Cola, Unilever, Heineken, Cargill, and Kraft Heinz (Invest in Holland, n.d.). As an agriculture powerhouse, the Netherlands could provide some instructive insights from the evaluation frameworks and/or smart practices in the agriculture, agri-food, and agri-tech industries for the IAF, which also delivers agri-innovation and agri-tech projects.

Results

After an extensive search on Dutch government websites, the research found two key Dutch governmental institutions “relevant to understanding evaluation” (Leeuw, 2009). The first is the National Audit Office (Algemene Rekenkamer), which is an independent body responsible for conducting audits of government agencies and programs. The second is the Ministry of Finance’s Inspectorate of Public Finances, which reviews budget proposals from ministries to measure their efficiency and effectiveness (Leeuw, 2009). The English website of the Rekenkamer

yielded several results for evaluations conducted/commissioned by them on various public policies in the last 20 years but the search could not find an evaluation framework, criteria, handbook, or guidelines for "smart" or "best" practices in policy evaluation, let alone policy evaluation in the agriculture sector. Similarly, the search function on the Dutch Ministry of Finance website did not return any positive results, probably indicating that the policy handbook and/or guidelines (if any) are likely only available in the Dutch language. An example is the 2016 report in Dutch titled "Evaluatievermogen bij beleidsdepartementen, praktijken rond uitvoering en gebruik van ex post beleids en wetsevaluaties" that provides an overview of evaluation institutions, programming, budgeting, evaluation process and finally, evaluation use in central Dutch policy departments (Haarhuis, 2020).

The Ministry of Foreign Affairs, however, published the "Evaluation policy and guidelines for evaluations" in 2009, which describes the evaluation policy and guidelines for evaluations conducted by the Policy and Operations Evaluation Department (IOB) of the Dutch Ministry of Foreign Affairs and "places the evaluation policy in the context of" the government-wide framework for the evaluation of government policy in general. IOB is a member of the Organization for Economic Cooperation and Development (OECD) and Development Assistance Committee (DAC) countries' network of evaluation services and also takes part in consultations between the evaluation services of the European Union, and of the countries united in the Nordic+ group (Ministry of Foreign Affairs, the Netherlands, 2009). As such, this 2009 document by IOB is likely only one of the few official government documents in English that describes the Dutch approach to policy evaluation, which is the focus of this scan. The document also provides a relatively detailed overview of program evaluation in government-affiliated development assistance related projects. Keeping this in mind, the key themes of the document are summarized.

Evaluation Policy and Guidelines for Evaluations (2009):

The guidelines define evaluation as "an assessment, as systematic and objective as possible, of an ongoing or completed project, program or policy, its design, implementation, and results… An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors" (p. 15).

IOB conducts both policy reviews (policy evaluation) and ex-post impact studies (impact evaluations). IOB impact evaluations generally focus on interventions (ranging from programs to various forms of budget support), whereas policy evaluations address themes or sectors (p. 16).

The evaluation criteria focus on five themes: efficiency, effectiveness, impact, relevance, and sustainability. These criteria are largely identical to those drawn up by the OECD and Development Assistance Committee (DAC) countries (p. 17).

Efficiency measures how (economically) the resources (or inputs) are applied and yield direct results. For example, the most economical or cost-effective input may not always be the most appropriate and there can be significant trade-offs between the quantity and quality of outputs.

Some questions to consider when evaluating efficiency are:

- Have appropriate inputs been deployed at the lowest possible cost?
- Have activities been conducted in an uncomplicated manner?
- Have decisions been made at the right level and bureaucratic hurdles avoided?
- Have overhead costs been kept to a minimum?
- Has duplication been avoided?
- Have conflicts during implementation been prevented or solved?
- Have outputs been achieved within the planned period and budget? (p. 17)

Effectiveness measures the extent to which the direct results of interventions (output) contribute to the sustainable achievement of policy objectives (outcome). An intervention is considered effective if its outputs have made a demonstrable contribution to achievement of the intervention's intended objectives. Efficiency and effectiveness do not always coincide: a positive assessment for one does not automatically imply a positive assessment for the other. Evaluating effectiveness aims to establish the causality, i.e., is the input generated output(s) with the desired effects? In some situations, it may be difficult to distinguish between correlation and causality in the absence of a counterfactual (what would have happened if the intervention had not occurred). The following steps can help evaluate for effectiveness:

- Measuring changes in the effect variables compared to the situation at baseline
- Finding/attributing causality of the observed changes to the intervention
- Assessing these changes vis-à-vis the program's objectives (p. 18).

Impact refers to the question that to what extent, the effects achieved (outcomes) have contributed to broader, objectives (which should be detailed in the evaluation process). While effectiveness looks at only the intended effects of an activity, impact refers to all possible economic, social, political, technical, and ecological effects at local, regional, or national level that have a direct or indirect effect on the target group or other parties (p. 19).

The evaluation process focuses largely on IOB's mandates and the dos and don'ts of the evaluation methodology of IOB related projects. There is not a lot of information on the use of logic models (mentioned as the "results chain"), the evaluation questions or the evaluation framework itself; although it claims to follow a "theory-based" evaluation process because

theory-based evaluation helps capture “what has happened in the past” and what has been achieved after implementing a particular policy.

The evaluation methodology section provides a quick overview of methods and various data collections tools such as document and literature studies, interviews, direct observation and measurement, surveys, and case studies (p. 35). The final section discusses the reporting structure which is also specific to IOB and their stakeholder needs.

Overall, the report can be read as a guiding document on how to commission a program evaluation for IOB projects, instead of a handbook on program evaluation.

New Zealand:

New Zealand’s (NZ) economy is dependent on agricultural and horticultural exports which comprise of 85-90% of its dairy, meat, fruit, and vegetable production (International Trade Administration, 2021). Its food and fiber export revenue forecasts were CAD 41.8 billion by 30 June 2022. The food and fiber sector forms 81.8% of its total exports, 11.1% of its GDP and employs 13.8% of its workforce, which makes NZ a major agriculture sector player globally.

New Zealand, therefore, can provide some valuable insights about evaluation policies and frameworks for this research (New Zealand Government, 2022).

Results:

The scan used many of the same terms and phrases that were employed in the jurisdictional scan of the Netherlands but it also did not yield agriculture, agri-food, or agri-tech sector-specific evaluation frameworks or guidelines. In their absence, the search expanded to include other policy evaluation toolkits and frameworks published by the NZ government agencies that provide some guidelines (with varying levels of details) on program evaluation in the social sector. For brevity and time constraints, one of the toolkits that provides the most recent, relevant, and detailed guide to the evaluation process in the social sector is summarized.

Making Sense of Evaluation: A Handbook for Everyone (2017)

The handbook, *Making sense of evaluation: A handbook for everyone***,** was developed by the Social Policy Evaluation and Research Unit (Superu), which is a government agency that generates evidence-based policy options to understand and solve “complex social issues.” The handbook specifically relates to the New Zealand social sector but provides a comprehensive overview of evaluation concepts and processes that can be useful for other organizations working in different sectors and countries. It comprises of four modules.

Module 1 discusses what evaluation is, including its different uses (such as formative & summative evaluations), and evaluation types (such as process, outcome, impact, developmental, and realist evaluations), and how to prepare for evaluation (p. 4) It also highlights that since funder requirements can change suddenly, a good monitoring system and evaluation strategy can help deal with those sudden changes with minimum effort, rather than having to change the entire system whenever requirements are altered (p. 6).

Module 2 discusses theory of change, logic models, and how the theory and models can improve programs. It provides knowledge about how to develop a logic model, whom to involve and how to engage stakeholders. It also shares skills, tools and techniques that can help when developing a logic model (p. 12). The key highlights of this module include:

- **Theory of change:** A comprehensive description of how and why a desired change is expected to happen in a particular context. It tries to understand the causal links between what you do and what change occurs – why you believe it will happen (p.4).
- **Logic model:** A diagram or picture of your theory of change that shows in simple terms how what you are doing is expected to lead to the changes you intend (p. 15). The key elements of the logic model are described below.
    - **Inputs** are the resources being used for the program. This includes money, people, knowledge, facilities, equipment, time, contributions in kind and a range of other things (p. 17).
    - **Activities** are a process that transforms these inputs into something else. Activities vary depending on program type, organizational mandates, funding eligibility etc (*ibid*).
    - **Outputs:** are the products(s) of program activities. Outputs should be countable (e.g., people mentored, people completing training courses, different types of social work provided) (*ibid*).
    - **Outcomes** are the changes that happen as a result of the activities. Outcomes can be divided into short-, medium- or long-term depending on project goals, the length of the project, and other considerations. Depending on the objectives of your program, you might have a reduction in youth offending because of the mentoring program, people returning to the labor force, improved parenting skills, people in warm, dry houses and so forth. It helps understand the change(s) the program was designed for, the change(s) that happened and when, and what change(s) didn't happen (and why). These changes can be reflected in the Outcomes Framework, which a tool that helps link what you want to achieve and how you will measure success (p. 17).

Figure 9: Example of a logic model (p. 21)

Module 3 describes measuring change- the things that need to be in place to know if the activities have made a difference. It details what is required to be able to *measure*, what we need to be able to *judge*, and of these, which are the most *important* (i.e., key performance indicators or KPIs). KPIs can be developed by answering the question: *What do you want to measure?* While many indicators can be potentially measured, not all of them may be relevant to program goals and objectives. KPIs narrow down and focus on the most important measures for program success (p. 26). The figure below shows some of the different types of questions and their corresponding indicators.

What are we looking to measure?	Type of indicator
What was the situation at the beginning of the process – where are you starting from? You need to know this to identify change.	**Baseline**
How much have you done? You need this information for management reporting, and to work out how efficiently you are carrying out your initiative.	**Output**
How much is this all costing? How do the costs break down? This is needed for good management, to report to funders and to work out efficiency.	**Financial**
Whether the processes are working well. Measuring efficiency will help you understand some things about your process, for example, waiting times, bottlenecks, etc.	**Process**
How big is the change that you have achieved? Are you on track to achieve your longer term objectives?	**Outcome**
What factors outside your control might have a significant influence on your outcomes? You might be able to identify some of these, but some may come out of the blue.	**Contextual**
What was likely to have happened if you had done nothing. This might be through comparing your clients to a group of people who are not part of your initiative or by other forms of comparative analysis. It is challenging to do and sometimes has moral/ethical implications, but needs addressing.	**Counterfactual**
How much of any change is a result of what you did? Much as you might like to claim all the credit, it is unlikely that you are the sole factor leading to success.	**Attribution**
The long term, big picture change at a population, society or place level.	**Impact**
The value of what has been achieved in relation to its cost. Both the cost and the value may be more than just financial costs or benefits. Crudely put, how much "bang for your buck"? Funders choosing between options will generally pick the one with the higher return.	**Economic impact/ Return on investment**

Figure 10 : Sample evaluation questions (p. 33)

A useful distinction to make while drafting performance indicators is between *wanting* to know (i.e., something is interesting, but not completely relevant) and *needing* to know (i.e., it is important for the program/funding/policymaking, etc). The indicators should be firmly tied back to the questions. Moreover, some indictors can be measured quite quickly to see if the program is on track, but others will not be noticeable until further down the road. For example, if a program goal is to reduce reoffending, and the measure is the number of people who do not reoffend within a 12-month period, the evaluators will need to wait for 12 months (p. 35). So, it is also important not to expect immediate results from long-term strategies for large-scale problems.

Module 4 discusses how to go about planning, commissioning, and managing evaluation. The evaluation plan can be developed by someone with research or evaluation experience, but it must have inputs from program staff involved in designing and implementing the program as well as buy-in from key stakeholders (p. 34). The key main things you need to cover in the plan are:

- Program outline: what needs to be evaluated
- Purpose and scope of the evaluation
- Key evaluation questions
- Timeline
- Available resources
- Risk management
- Outcomes framework and associated indicators (p. 35).

Figure 11 below provides some example of key evaluation questions. It should not, however, be considered exhaustive and should be tailored to match program evaluation needs.

If you are evaluating the outcomes of the programme, you will be asking questions such as:

- How well did the programme work?
- Did the programme produce or contribute to the intended outcomes in the short-, medium- and long-term?
- For whom, in what ways and in what circumstances were outcomes produced?
- What unintended outcomes (positive and negative) were produced?
- To what extent can changes be attributed to the programme?
- What were the particular features of the programme and context that made a difference?
- What was the influence of other factors?
- What innovations does the programme present?

If you are evaluating how the programme is working, you will be asking questions such as:

- How is the programme being implemented?
- How appropriate are the processes compared with quality standards?
- Is the programme being implemented correctly?
- Are participants being reached as intended?
- How satisfied are different groups of clients?
- What has been done in an innovative way?
- What, if any, things have happened that were unexpected or affect other people?

Figure 11: Example of key evaluation questions (p. 37)

Finally, allocating a budget for program evaluation is a key consideration, especially for nonprofits, which often have limited funding and resources and competing demands from the

funding agencies. A budget can depend on several factors such as program goals and its scale – a complex program with multiple goals would likely require more resources, program length- a new or pilot program would likely need more time and resources (and a bigger budget) to measure what, when and why, compared to an existing program that's been running for years which might already have significant available data. The evaluation budget can also vary depending on who conducts the evaluation- eternal evaluators would likely cost more compared to internal evaluators (pp. 40-41).

Apart from this handbook, the jurisdictional scan found a few other publications related to monitoring and evaluation published by the New Zealand public agencies. Two of the most relevant of these publications are Guidance Note: Best Practice Monitoring, Evaluation and Review (2019) by the New Zealand Treasury and the Evaluation Operational Policy (2020) published by the New Zealand Foreign Affairs and Trade Aid Program.

This Guidance Note describes the key elements of Monitoring, Evaluation and Review. These elements should underlie the development of all government policy initiatives that involve proposals to create, amend or repeal primary or secondary legislation (a 'government regulatory proposal') (New Zealand Government, 2019). The Note is 10 pages long and unlike the Making Sense of Evaluation Handbook, which explains the various aspects of the evaluation program in detail, it only provides a quick overview of the evaluation process. The themes explored are: what is evaluation? The purpose of monitoring and evaluation, logic models and how to incorporate them in the policy process.

The Evaluation Operational Policy, on the other hand, sets out the roles, responsibilities, expectations, and standards for evaluations funded by New Zealand's Official Development Assistance in Pacific Island countries. It provides a high-level overview of the evaluation process (similar to the Guidance Note) and is specifically targeted at the New Zealand government-affiliated entities working in these countries.

Most of the information in the two documents is about evaluation, evaluation types (e.g., summative, and formative evaluations) and logic models, which has already been covered in the Handbook and earlier sections of this research. Apart from these toolkits, the Treasury NZ (2021) has developed a Cost-benefit Analysis (CBA) tool called CBAx to ensure that robust value for monetary assessment is applied to investment and budget decisions. CBAx helps agencies conduct in-house CBAs, with a focus on i) a set of common assumptions when quantifying and monetizing the impacts of different proposals (for example, impact values and discount rates); ii) having a robust framework for estimating the broader societal impacts of policy options; and iii) having New Zealand-specific publicly available information to value impacts. CBAx analysis can provide an overall benefit cost ratio, return on investment (RoI) and a net present value (NPV) for a proposal. It can also prompt clarity about the intervention logic, the counterfactual, the impact assumptions about who is affected, in what way, to what extent, when and for how long, and the evidence base (New Zealand Treasury 2021, pp. 6-9).

On the other hand, the CBAs of agricultural research present challenges that other types of impact assessments do not, such as dynamic biological environments, complex farming systems, aggregation issues and difficulties in identifying appropriate counterfactuals. Therefore, it is "not possible" to use the CBAx to measure returns to agricultural research in terms of costs and benefits (Yangab & Zydenbosa, 2021). Given these constraints, the detailed summary of the CBAx tool was excluded from this jurisdictional scan.

Summary of Jurisdiction Scan Findings

The Netherlands and New Zealand (NZ) both have a large agricultural economy. The Dutch agricultural sector exports some CAD 85 billion worth of agricultural produce annually whereas New Zealand's food and fiber sector's annual exports were forecast at CAD 41.8 billion by June 30, 2022. As such, IAF identified that these countries could provide some instructive insights from the evaluation frameworks and/or smart practices in the agriculture, agri-food, and agri-tech industries.

One of the key findings was that there is a lack of evaluation frameworks specific to the agriculture or agri-tech or agri-food sector, in general. In the Netherlands particularly, this issue was compounded because the search could not find publicly available evaluation frameworks in English, except for their Ministry of Foreign Affairs' "Evaluation policy and guidelines for evaluations (2009)". Therefore, the search expanded to include evaluation frameworks in other comparable sectors such as the public and social sectors in both countries.

The Dutch document's evaluation criteria focus on five themes: efficiency, effectiveness, impact, relevance, and sustainability. The evaluation design provides a quick overview of methods and various data collections tools such as file and literature studies, interviews, direct observation and measurement, surveys, and case studies (p. 35) but the report reads more as a guiding document on how to commission a program evaluation than a handbook on program evaluation.

For New Zealand, "Making sense of evaluation: A handbook for everyone (2017)"**,** published by the government's Social Policy Evaluation and Research Unit provides a comprehensive overview of evaluation concepts and processes for the social sector that can be useful for other countries and sectors. The NZ document is divided into four modules: i) Module 1 discusses what evaluation is, including its different uses (such as formative & summative evaluations) (p. 4); ii) Module 2 discusses theory of change and logic models and how they can improve programs; iii) Module 3 describes measuring change and provides an overview of the most important measurement indicators or the key performance indicators; and iv) Module 4 discusses how to go about planning, commissioning and managing evaluation.

For NZ, the search also analyzed other government documents such as Guidance Note: Best Practice Monitoring, Evaluation and Review (2019) by the New Zealand Treasury and the Evaluation Operational Policy (2020) published by the New Zealand Foreign Affairs and Trade

Aid Program but these only contain a cursory overview of the evaluation design and process, so the 2017 Handbook was the most comprehensive resource.

Chapter 7: Discussion and Analysis

This chapter discusses and analyses the key findings from interviews, jurisdictional scan, and the document review. The main themes from the findings highlight the state of existing evaluation practices at IAF and discuss the key features of the proposed evaluation framework in the light of findings from the literature review and the jurisdictional scan. These collectively help answer the main research question: What makes for an effective evaluation framework for IAF? This chapter also discusses the limitations of this study and raises some new research questions.

Main Themes in Research

Theme 1 - IAF evaluations tend to be accountability-focused

One of the main themes that emerged during the interviews is that current evaluation practices at IAF are largely accountability-based and are focused on keeping track of eligible program activities and total funds spent on those activities. This finding is corroborated by the document review of IAF's financial reporting templates (which document how the project money is spent), and the project activities reporting templates (which track the activities undertaken with the funds).

This finding is not surprising and in fact, it is consistent with the literature on program evaluation for nonprofits, especially those working in the agriculture sector, which suggests that most donor-driven programs tend to be accountability-based (Chapter 2: Literature Review, p. 11). The reason for the increased emphasis on accountability is because of the commercialization of the nonprofit sector and the competition between for-profits and nonprofits for funding and resources and consequently, the need to be accountable for both (Salmon, 2016; Literature Review, p. 9). With an increasing number of nonprofits competing for the delivery of public goods and services, the focus on accountability of donor funds likely gives IAF a competitive advantage over other nonprofits in the sector. Accountability can also be a useful way for nonprofits to manifest, exert and maintain their independence vis-a-vis their mission (Carman 2009, p. 375; Chapter 2: Literature Review, P. 10).

Lastly, accountability can be considered the raison d'être for IAF to commission this study: develop a program evaluation framework that demonstrates value-for-money to their funding partners.

Theme 2 - IAF needs to conduct more field visits to monitor and measure program impacts

Another theme that emerged during the interviews was the lack of on-field program monitoring by IAF. Instead, it seems, for most programs, IAF mostly relies on client-supplied information (such as invoices for the purchases, receipts for trade shows, video of advertising/marketing campaign etc.). The interviews highlighted that not all programs require filed visits but some,

especially those, with an on-farm component, could benefit from them. Another reason behind less on-field monitoring is capacity constraints (i.e., lack of staff time and resources). It requires dedicated staff time and fiscal resources to conduct field monitoring, which nonprofits often struggle with, in the face of competing priorities. Therefore, IAF should consider spending time and resources on establishing effective monitoring mechanisms because as discussed in the Literature Review (p. 9), irrespective of the evaluation type, monitoring helps ensure the programs are implemented in accordance with their design and objectives as it helps answer questions such as "Are we doing the right thing? Are we doing it right?" (Suvedi & Stoep 2016, p. 3).

Monitoring tracks inputs, activities, outputs, outcomes and impacts throughout the project and yields valuable data for either course-correction or future improvements about the project. (World Bank, n.d.; Chapter 2: Literature Review, p. 9) Moreover, it is also supported by the jurisdictional scan findings (p. 32) that having a good monitoring system in place is particularly useful for nonprofits who rely on donor-driven projects. Since funder requirements can often change suddenly, a good monitoring system and evaluation strategy can help deal with those sudden changes with minimum effort, rather than having to change the entire system whenever requirements are altered.

Theme 3 - IAF program goals need to be more targeted

Another key finding from the interviews was that most programs have very broad objectives and therefore IAF receives a lot of applications for them, most of which are ineligible (Chapter 4: Interview Findings, p. 25). For example, last year, IAF had to turn down around 90% of applicants after reviewing applications because they could only select the top applications based on their selection criteria (the specific selection criteria was not shared). This was identified as not only a waste of time for the people who applied and got rejected but also an inefficient use of IAF's staff time. To avoid that, IAF needs more targeted programs that mostly attracts relevant applications. This may be beyond IAF's control, because as described in the interviews (p. 22), program goals and objectives are often set by the funding partners.

Despite these constraints, there are some programs in which IAF plays a developmental role that involves developing program logic models to outline program goals, inputs, activities, and outputs. Going forward, IAF is looking at utilizing and repurposing some of the left-over funding from their existing projects to develop more of these programs (Chapter 4: Interview Findings, p. 27). This could be an opportunity for them to develop programs with fewer, targeted, and measurable objectives that also tie in with their organizational goals. Having targeted goals also makes it easier to develop key performance indicators (KPIs) to measure if program activities are achieving the desired program goals. The KPIs are the most important measures that need to be in place to know if the activities have made a difference (Chapter 6: Jurisdictional Scan Findings, p. 34). KPIs also help identify areas of improvement and other barriers affecting program performance (United States Environmental Protection Agency, 2022; Chapter 2: Literature

Review, p. 10). For instance, one of the KPIs that frequently came up during the interviews was the unavailability of baseline data- data collected prior to intervention or program (Superu, 2017, p. 3)- to compare program impacts before and after a program was implemented (Chapter 4; Interview Findings, p. 27).

It is also important to note that measuring change also requires sufficient timeline to capture some of the program impacts that depending on a project, could take anywhere between a few months to a few years after project completion. Nonprofits including IAF, often have competing priorities and capacity constraints to revisit and evaluate program impacts after six months, a year or two years after program completion (Chapter 4: Interview Findings, p. 27).

Theme 4 - IAF needs to integrate formative evaluation components in their evaluation practices

The interviews also highlighted the need to integrate formative evaluation components in IAF's evaluation practices (Chapter 4: Interview Findings, p. 25). As described in the Literature Review (p. 8), a formative evaluation is a proactive endeavor to develop a program or policy, anticipate and rectify its potential challenges and to measure and record key performance indicators of a policy or program (Alkin & Vo, 2018, p. 13). Presently, IAF's evaluation practices are summative, especially for their sales-related programs (Chapter 4 Interview Findings, p. 2). Summative or outcome evaluations are conducted towards the end of the program to provide decisive input about the success or failure of the program activities vis-à-vis program objectives (Alkin & Vo, 2018, p. 13). On the other hand, formative or process evaluations, are often conducted in the early stages of a program to identify any challenges and determine course-correction (Alkin & Vo, 2018, p. 13).

This emphasis on formative evaluations not only complements the other key finding from the interviews about the need to develop targeted programs but is, by definition, necessary for it. As IAF looks to design, develop, and implement its own programs in the future from repurposed funding, asking questions about how the program will be delivered, for whom, why, and under what conditions- key features of a formative evaluation- would become invaluable.

Theme 5 - IAF should integrate data visualization software in their evaluation practices

Another main theme that emerged during the interviews was integrating modern software, such as Power BI or Tableau in IAF's program evaluation practices (Chapter 4: Interview Findings p. 26). These are interactive data visualization software with a primary focus on business intelligence and can act as internal dashboards to provide real-time reporting on various program activities: such as keeping track of program budgets activities for better monitoring and evaluation. These dashboards can be useful in identifying program weaknesses or challenges in program delivery that can be relatively quickly remedied, which would otherwise have to wait till project completion. The software can be very useful in terms of communicating within, and also to quickly tell a story to the funding partners or provide information to IAF's public

partners. The focus on tech-based solutions is part of ongoing efforts by IAF to modernize and improve its existing project reporting, evaluation, and communications infrastructure. For example, IAF moved from paper-based applications files for its programs to online forms (Chapter 4: Interview Findings, p. 25). The program surveys and reporting templates are also either sent as online forms and/or as MS Office files through email. According to IAF, this has helped improve efficiency of program delivery (Chapter 4: Interview Findings, p. 24). Since most of IAF's clients are small farmers who may not be as tech-savvy, IAF also needs to strike a balance between continuously improving efficiency of their program delivery and their program recipients' needs.

Chapter 8: Conclusions and Recommendations

This thesis aimed to explore effective program evaluation frameworks, particularly in the agri-sector, for monitoring and evaluation of public and private sector projects focusing on the agriculture, agri-food, and agri-tech sectors to answer the main research question: What is the most effective program evaluation framework for Investment Agriculture Foundation (IAF)? IAF is acting as an informal client for this thesis to provide a pragmatic approach to research, which is consistent with the discipline of public administration.

The need for this research was identified by IAF as their current program evaluation process largely consists of summarizing the amount of funding delivered to stakeholders and collecting program-specific reporting metrics. They are looking draw on smart evaluation practices and frameworks, particularly in the agrisector sector (where possible) that helps them assess program effectiveness and impacts. In doing so, the thesis intended to contribute to the literature on effective evaluation practices and frameworks for programs that fund and support agriculture, agri-food, and agri-tech initiatives, and how these practices/frameworks can be adapted to different geographic and organizational contexts.

The research followed a four-pronged approach, to find answers to the main research questions, comprising of a) literature review, b) IAF staff interviews, c) jurisdictional scan and d) document review. The literature review introduced, defined, and discussed key themes such as program evaluation, the relationship between evaluation and accountability, and monitoring and evaluation, and the similarities and differences between evaluation and performance measurement. It also provided the most recent and relevant literature on the challenges of conducting program evaluations in the nonprofit sector, particularly nonprofits that work in the agriculture sector.

The interviews with IAF staff, jurisdictional scan of program evaluation practices and frameworks in comparable economies such as the Netherlands and New Zealand, and the document review of existing IAF evaluation and reporting templates highlighted evaluation frameworks (where possible) and practices used by both the public and private sector for IAF's consideration. Some of the key conclusions for IAF, going forward, include i) developing more targeted programs with clearly outlined goals that are measurable through key performance indicators; ii) incorporating formative evaluation components in their evaluation process (i.e., asking questions about how the program will be delivered, for whom, why, and under what conditions); and iii) using data visualization software for improved and real-time monitoring of their programs.

One of the implications of the research, particularly from the literature review and jurisdictional scan, was the understanding that turn-key evaluation frameworks in the agriculture, agri-sector and/or agri-tech sectors in public or private sectors that can be readily adapted by IAF are either unavailable or publicly inaccessible. There are nonprofits such as AgResults, a $152 million initiative by the G20 countries, that have come up with innovative programs such as Pay-for-Results (PfR) prize competitions to incentivize the private sector to invest in high-impact agricultural innovations (AgResults, n.d.; Chapter 2: Literature Review, p. 12). Apart from sharing their overall evaluation approach, most of the information related to AgResults' evaluation frameworks either seems proprietary, as the organization hires external evaluators for their program evaluations and/or not publicly accessible.

Despite these limitations, this research has analyzed and identified a few recent and relevant program evaluation frameworks and practices such as the New Zealand government-published "Making Sense of Evaluation: A Handbook for Everyone (2017)" and the Dutch Ministry of Foreign Affairs'-published "Evaluation Policy and Guidelines for Evaluations (2009)". These documents provide some valuable insights about developing a program evaluation framework for public and social sector programs, many of which can and should be applied at IAF.

Moreover, another key (and positive) implication of this research is the realization that the basic ingredients of an effective program evaluation framework already exist in IAF's evaluation practices that can be further refined. For instance, one of the main interview findings was that IAF's evaluation practices are accountability-based, demonstrated by their tracking of projects funds and activities through their financial and activity reporting templates, respectively, and reporting it to their funders.

Accountability is both a sought-after quality in donor-funded projects that likely gives IAF a competitive advantage over other nonprofits in the sector, and a key element of program evaluations. This is complemented by the practice of developing logic models for programs in which IAF plays a developmental role such as the Buy BC program and CAP Market development programs. A logic model is one of the most important parts of any evaluation process that shows in simple terms how what you are doing is expected to lead to the changes you intend. These are just of the main related findings to emerge from this research.

Recommendations

The main recommendations from the research findings to support an improved program evaluation framework for IAF are the following:

1. **Encourage discussion and collaboration with potential funders at the outset of the project to develop more targeted programs with clearly outlined goals that are measurable through key performance indicators**.

This may not be possible for those programs in which program objectives are predetermined by the donors.

2. **Incorporate formative evaluation components in IAF's evaluation process** (i.e., asking questions about how the program will be delivered, for whom, why, and under what conditions), alongside summative components (measuring the success or failure of the program activities vis-à-vis program objectives); which are already part of its evaluation practices, especially for its sales and marketing programs.

3. **Consider adopting data visualization software such as Tableau or Power BI for improved and real-time monitoring of their programs.**

4. **Engage a researcher to build on the learnings from this research towards building a more effective program evaluation framework**.

This is a longer engagement (at least 3-4 months) and there are some costs associated with this approach such as IAF staff and student's time spent on it.

Moving Forward

This research and its findings should be considered a stepping-stone for further engagement with and within IAF to build on and refine existing evaluation practices using evaluation insights from comparable sectors and jurisdictions, and more importantly, the knowledge and learnings from the interviews, towards building a fulsome program evaluation framework working with its stakeholders to determine their needs and preferences are being met.

In addition to IAF, this research should also be useful for the broader literature on effective evaluation practices and frameworks for programs in the nonprofit sector and those that fund and support agriculture sector initiatives.

References

AgResults. (n.d.). *What is AgResults?* Retrieved from AgResults: https://agresults.org/our-approach

Agriculture and Agri-Food Canada. (2019). *Evaluation of the Canadian Agricultural Adaptation Program.*

Agriculture and Agri-Food Canada. (2021). *Evaluation of the programs under the Agricultural Marketing Programs Act (2014 to 2019).*

Agri-Food Economic Strategy Table. (2017). *The Innovation and Competitiveness Imperative: Seizing Opportunities for Growth Report of Canada's Economic Strategy Tables: Agri-food.*

Alkin, M., & Klein, S. (1972). Accountability Defined: Evaluating Teachers for Outcome Accountability. *UCLA Evaluation Comment.* 3(1), 1-11. Retreived from https://files.eric.ed.gov/fulltext/ED068495.pdf

Alkin, M., & Vo, A. (2018). *Evaluation essentials: From A to Z.* Guilford Press.

Anderson, J. R., & Feder, G. (2004). Agricultural Extension: Good Intentions and Hard Realities. *Oxford University Press*, *19*(1), 41-60.

Carman, J. (2009, July). Nonprofits, Funders, and Evaluation Accountability in Action. *The American Review of Public Administration, 39*(4), 374.

Carnochan, S., Samples, M., Myers, M., & Austin, M. (2013). Performance Measurement Challenges in Nonprofit Human Service Organizations. *Nonprofit and Voluntary Sector Quarterly*, 1-5.

Centers for Disease Control and Prevention. (1999). *Summary of the Framework for Program Evaluation.*

CSU Library. (2021, September 1). *Literature Review: Traditional or narrative literature reviews*. Retrieved from https://libguides.csu.edu.au/review/Traditional

Dauphinee, W. D. (2015). The role of theory-based outcome frameworks in program evaluation: Considering the case of contribution analysis. *Medical teacher*, *37*(11).

Davies, R. (2012). *AgResults: Innovation in Research and Delivery.*

Donaldson, S. I. (2021). Chapter: Developing Program Impact Theory. In S. I. Donaldson, *Introduction to Theory-Driven Program Evaluation* (p. 27). New York: Routledge.

FAO. (2020). *Review of monitoring and evaluation capacities in the agriculture sector – Study conducted in collaboration with EvalForward,a Community of Practice on Evaluation for Agriculture, Food Security and Rural Development.* Rome.

Flay, B. F., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science*, 1-24.

Fynn, J., Wendy, H., Milton, K., & Jones, A. (2020). A scoping review of evaluation frameworks and their applicability to real-world physical activity and dietary change programme evaluation. *BMC Public Health*, 1-15.

G20 Summit. (2010, June 26-27). *THE G-20 TORONTO SUMMIT DECLARATION- June 26-27, 2010.* Retrieved from G20- Toronto: http://www.g20.utoronto.ca/2010/g20_declaration_en.pdf

Government of British Columbia. (n.d.). *BC Agritech.* Retrieved from https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/agriculture-and-seafood/innovation-and-technology/7206_agritech_info_sheet.pdf

Government of Canada. (2021, December 08). *Project Managers' Guide to Performance Measurement and Evaluation*. Retrieved from https://www.justice.gc.ca/eng/rp-pr/cj-jp/fv-vf/pmgui-guigp/p3.html

Government of the Netherlands. (n.d.). *Agriculture and Horticulture*. Retrieved from Agriculture: https://www.government.nl/topics/agriculture/agriculture-and-horticulture

Grosvener Public Sector Advisory. (n.d.). *Program Theory and Program Logic: Articulating how your program works*. Retrieved from https://www.grosvenor.com.au/public-sector/program-evaluation/articulating-how-your-program-works/

Haarhuis, C. K. (2020). The Netherlands. In R. Stockmann, M. Wolfgang, L. Taube, & (Eds), *The Institutionalisation of Evaluation in Europe* (pp. 89-114). Cham: Palgrave.

International Trade Administration. (2021, August 10). *New Zealand - Country Commercial Guide: Agricultural Sector - Food and Beverages*. Retrieved from International Trade Administration: https://www.trade.gov/country-commercial-guides/new-zealand-agricultural-sector-food-and-beverages

Invest in Holland. (n.d.). *An Agrifood Powerhouse: Small Country with Global Reach Feeds the World*. Retrieved from Invest in holland: https://investinholland.com/doing-business-here/industries/agrifood/

IOB (2009). Evaluation Policy and Guidelines for Evaluations. Ministry of Foreign Affairs, the Netherlands. 1-30. Retrieved from: https://english.iob-evaluatie.nl/publications?page=3

James C. McDavid, J. C., & Hawthorn, L. R. (2010). *Key concepts and issues in porgram evaluation and performance measurement.* Sagepub.

Kuada, J. (2012). *Research methodology: - A Project Guide for University Students.* Frederiksberg: Samfundslitteratur.

Lee, C., & Nowell, B. (2015). A Framework for Assessing the Performance of Nonprofit Organizations. . *36*(3), 299-319.

Leeuw, F. (2009). Evaluation policy in the Netherlands. In W. Trochim, M. Mark, & L. (. Cooksy, *Evaluation policy and evaluation practice. New Directions for Evaluation* (pp. 87–102).

Leeuw, F. (2003). Reconstructing Program Theories: Methods Available and Problems to be Solved. *American Journal of Evaluation, 24*(1), 5-20.

Martin, P. (2019). The Logic Model: A look at Outcomes and Indicators. Better Strategies Retrieved from https://www.brighterstrategies.com/blog/the-logic-model-a-look-at-outcomes-and-indicators/

Mertens, D. M., & Wilson, A. (2019). *Program Evaluation Theory and Practice, Second Edition : A Comprehensive Guide.* New York: Guilford Press.

Ministry of Foreign Affairs, the Netherlands. (2009). *Evaluation policy and guidelines for evaluations.*

Mitchell, G. E., & & Berlan, D. (2018). Evaluation in Nonprofit Organizations: An Empirical Analysis. *Public Performance & Management Review , 41*(38), 416.

Moullin, M. (2002), Delivering Excellence in Health and Social Care, Open University Press, Buckingham.

Narayan, T. A., Geyer, J., Mainville, D., & Ness-Edelstein, B. (2022). An evaluation framework and empirical evidence on the effect of pay-for-results programs on the development of markets for welfare-enhancing agricultural technologies. *Agriculture & Food Security, 11*.

New Zealand Government. (2022). *Situation and Outlook for Primary industries.* Wellington: Ministry of Primary Industries.

New Zealand Treasury. (2021). *CBAx Tool User Guidance.* New Zealand Government.

OVC. (2010, January). *What Are Performance Measurement and Program Evaluation?* Retrieved from Guide to Performance Management and Program Evaluation: https://www.ovcttac.gov/docs/resources/OVCTAGuides/PerformanceMeasurement/whatarepm_pfv.html

Patton, M. (2008). *Utilization-focused Evaluation, 4th Edition.* Thousand Oaks: Sage.

Peer, N. (2017). Evaluating Cultural Competency Concepts Using Program Theory-Driven Evaluation Framework. *Nursing education perspectives, 38*(6), 337–339.

Public Health Ontario. (2016). *Logic model–A planning and evaluation tool.*

Rogers, P. (n.d.). *Develop programme theory / theory of change*. Retrieved from Better Evaluation: https://www.betterevaluation.org/en/rainbow_framework/define/develop_programme_theory

Rowan, M., Hogg, W. L., Kristjansson, E., & Dahrouge, S. (2009). A Thoery-based Evaluation Framework for primary Care: Setting the Stage to Evaluate the "Comparison of Models of Primary Health in Ontario" Project. *The Canadian Journal of Evaluation*, 113-140.

Salamon, L. M. (2016). *The Resilient Sector: The Future of Nonprofit America.* New York: Brookings.

Sharpe, G. (2011, November). A Review of Program Theory and Theory-Based Evaluations. *American International Journal of Contemporary Research, 1*(3), 72-75.

Stone-Jovicich, S., Percy, H., White, T., Turner, J. A., & Chen, L. (2019). Evaluating monitoring, evaluation and learning initiatives in the New Zealand and Australian agricultural research and innovation systems: The MEL2 framework. *Evaluation Journal of Australasia*, 8-19.

Superu (2017). Making Sense of Evaluation: A handbook for everyone. Government of New Zealand. 1-46.

Suvedi, M., & Stoep, G. V. (2016). *Improving the Monitoring and Evaluation of Agricultural Extension Programs.* MEAS Discussion Paper Series on Good Practices and Best Fit Approaches in Extension and Advisory Service Provision.

Tatian, P. A. (2016, March). *Performance Measurement to Evaluation.* Retrieved from Urban Institute: https://www.urban.org/sites/default/files/publication/78571/2000555-performance-measurement-to-evaluation-march-2016-update_2.pdf

Taylor, A., & Liadsky, B. (2016). *Evaluation Literature Review.* Toronto: Ontario Nonprofit Network.

Treasury Board of Canada Secretariat. (2021, March 22). *Theory-Based Approaches to Evaluation: Concepts and Practices*. Retrieved from https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/theory-based-approaches-evaluation-concepts-practices.html

Treasury Board of Canada Secretariat. (2012). *Program Evaluation Methods: Measurement and Attribution of Program Results.*

USDA. (2022, June 22). *Risk in Agriculture*. Retrieved from Economic Research Service U.S. Department of Agriculture: https://www.ers.usda.gov/topics/farm-practices-management/risk-management/risk-in-agriculture

United States Environmental Protection Agency. (2022, March 25). *Program Evaluation and Performance Measurement at the EPA.* Retrieved from https://www.epa.gov/evaluate/program-evaluation-and-performance-measurement-epa

Verdung, E. (1997). *Public Policy and Program Evaluation.* (https://doi.org/10.4324/9781315127767, Ed.) Routledge.

Walk, M., & Kennedy, T. (2016, August). *Making Nonprofits more Effective: Performance Management and Performance Appraisals.* Retrieved from CORE: https://core.ac.uk/reader/84831680

Winters, P., Salazar, L., & Mafioli, A. (2010). *Designing Impact Evaluations for Agricultural Projects.*

World Bank. (n.d.). *Monitoring and Evaluation*. Retrieved from https://ieg.worldbankgroup.org/what-monitoring-and-evaluation

Yangab, W., & Zydenbosa, S. (2021). Ex-ante economic evaluation of agricultural research in New Zealand – a conceptual framework. Lincoln, New Zealand.

Appendices

Appendix A: Interview Invite Templates

Subject: Request to interview for IAF program evaluation framework study

Dear [FULL NAME],

I hope you are doing well.

I am Zulfiqar Ali, a second-year student in the Master of Public Administration program at the University of Victoria. I am working with Natalie Janssens, Director of Programs at Investment Agriculture Foundation (IAF), to develop the new program evaluation framework for IAF, as part of my master's thesis. The research is being conducted under the supervision of Dr. Kimberly Speers, Assistant Teaching Professor and Master of Public Administration Academic Coordinator at the University of Victoria.

The study deliverable will outline an approach for IAF to evaluate programs using tools such as: logic models, specific activity reporting metrics, and surveys to help track and measure inputs, activities, outputs, outcomes, and impacts, when bidding to deliver new government programs.

Your participation in a 45-minute interview will greatly help identify the challenges faced by IAF and their partners in existing program evaluation framework(s) and brainstorm ideas for the new program evaluation framework. A sample list of questions is attached for your reference. Please understand that your participation in the interview will be completely voluntary with the option to opt out of the interview at any time should you wish to do so.

In this regard, could you please read the attached consent form to participate in the interview and email me a signed copy by *[insert time and date]? A meeting link will be shared in a separate email after receiving the signed consent form.*

Please let me know if you have any questions.

Warm regards,

Zulfiqar

Here's the Zoom meeting link:

[Insert Zoom meeting link]

Follow Up Email Template (in case of no response)

Hello [insert name]

I hope you're doing well.

This is Zulfiqar Ali, a second-year student in the Master of Public Administration program at the University of Victoria. I am working with Natalie Janssens, Director of Programs at Investment Agriculture Foundation (IAF), to develop the new program evaluation framework for IAF, as part of my master's thesis. The research is being conducted under the supervision of Dr. Kimberly Speers, Assistant Teaching Professor and Master of Public Administration Academic Coordinator at the University of Victoria. **Please see my previous email for details about the study and the list of interview questions.**

This is a gentle reminder to secure your participation for the 45-minute interview will greatly help identify the challenges faced by IAF and their partners in existing program evaluation framework(s) and brainstorm ideas for the new program evaluation framework. Please understand that your participation in the interview will be completely voluntary with the option to opt out of the interview at any time should you wish to do so. More details about your participation can be found in the attached consent form.

Should you agree to participate in the interview, please email me a signed interview consent form by {Insert time and date}

Please let me know if you have any questions.

Warm regards,

Zulfiqar

Appendix B: Consent Form Template

University of Victoria

PARTICIPANT CONSENT FORM

Developing the Program Evaluation Framework for Investment Agriculture Foundation of British Columbia

You are invited to participate in the subject study that is being conducted by Dr. Kimberly Speers and Zulfiqar Ali.

Dr. Kimberly Speers is an Assistant Teaching Professor and Master of Public Administration Academic Coordinator at the University of Victoria and you may contact her if you have further questions by kpseers@uvic.ca.

As a Graduate student, I am required to conduct research as part of the requirements for a Master's degree in Public Administration. It is being conducted under the supervision of Dr. Kimberly Speers.

The IAF staff (particularly Natalie Janssens) is the informal client for this study. Some of the IAF staff, such as yourself, are also being requested to participate in the interviews to share institutional knowledge and key insights about developing the new program evaluation framework. Your input is critical in understanding the challenges in existing program evaluation efforts and in identifying ways to remedy those challenges in the new program evaluation framework.

This research is being funded by Investment Agriculture Foundation.

Purpose and Objectives

Currently, Investment Agriculture Foundation's program evaluation activities are varied and based on a funding agency's requirements for each program that the organization is contracted to deliver. The evaluation requirements are a summary of funding delivered to stakeholders as opposed to an assessment of the program's efficacy and impact. The Foundation's main funding partners (government departments & ministries) have expressed a desire to demonstrate the value for money and the impact of programs funded through taxpayer dollars. In this regard, the Master's Thesis' main deliverable will be a standalone document that outlines an approach for IAF to evaluate programs using tools such as: logic models, specific activity reporting metrics, and surveys (as well as any others identified) to help track and measure inputs, activities, outputs, outcomes, and impacts.

Importance of this Research

The Master's thesis' main deliverable will be a standalone document that outlines an approach for IAF to evaluate programs using tools such as: logic models, specific activity reporting metrics, and surveys (as well as any others identified) to help track and measure inputs, activities, outputs, outcomes, and impacts.

Participants Selection

You are being asked to participate in this study because of your experience and knowledge of administering/evaluating IAF programs.

What is involved

If you consent to voluntarily participate in this research, your participation will include audio/video-recordings of the interview (approximately 45 minutes) and written notes. A transcription will be made.

Inconvenience

Participation in this study may cause some inconvenience to you, including taking up more than 45 minutes of your time (if needed).

Risks

There are no known or anticipated risks to you by participating in this research.

Benefits

The potential benefits of your participation in this research include increased knowledge to develop an effective program evaluation framework for your organization.

Voluntary Participation

Your participation in this research is completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw from the study your data will only be used upon your consent.

Anonymity

Given the small group of potential participants and the purposive sampling taking place, it will be impossible for those who do or do not participate in the study to remain fully anonymous. If you have any questions or concerns about the use of your information, please let me know.

Confidentiality

Your confidentiality and the confidentiality of the data will be protected by securing all interview notes electronically by the Principal Applicant.

Dissemination of Results

It is anticipated that the results of this study will be shared with others in the following ways: thesis/dissertation, presentations at scholarly meetings and IAF meetings.

Disposal of Data

Data from this study (only saved electronically) will be erased after the thesis is deemed "pass" by the thesis committee and IAF has received their final deliverables.

Future use of data:

The Investment Agriculture Foundation may utilize the information gathered during this interview for future uses.

Contacts

Individuals that may be contacted regarding this study include Zulfiqar Ali (Principal Applicant), Dr. Kimberly Speers (Principal Investigator) and Natalie Janssens (Director of Programs).

In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Office at the University of Victoria (250-472-4545 or ethics@uvic.ca).

Your signature below indicates that you understand the above conditions of participation in this study, that you have had the opportunity to have your questions answered by the researchers, and that you consent to participate in this research project.

_____	_____	_____
Name of Participant	*Signature*	*Date*

PLEASE SELECT STATEMENT only if you consent:

I consent to be identified by name / credited in the results of the study: _____ (Participant to provide initials)

I consent to have my responses attributed to me by name in the results: _____ (Participant to provide initials)

Future Use of Data *PLEASE SELECT STATEMENT:*

I consent to the use of my data in future research: _____ (Participant to provide initials)

I **do not** consent to the use of my data in future research: _____ (Participant to provide initials)

I consent to be contacted in the event my data is requested for future research: _____ (Participant to provide initials)

A copy of this consent will be left with you, and a copy will be taken by the researcher.

Appendix C: Interview Questionnaire

1. Could you please briefly describe your position and role at IAF?
2. What are some of the key existing input and output indicators of success programs delivered by IAF?
    a. Which indicators are common among the programs, and which are program-specific? – Are there any supporting data/info/docs that can help me better understand the process?
    b. What are key data/information sources for the indicators?
    c. How were the indicator systems developed? Did IAF draw upon other indicator systems and frameworks?
3. To what extent are the inputs, activities, outputs, and outcomes in the evaluation framework(s) consistent with IAF program objectives?
    a. Are you aware if logic models or other tools were used in drafting the existing evaluation framework?
    b. Are participants being reached as intended? How are program clients contacted regarding the success of their projects? Is this an effective method of communication?
    c. Do you measure client satisfaction? If so, how? (i.e., what instruments do you use?)
    d. What areas (if any) have the program clients indicated dissatisfaction or desire for improvement?
4. What are the key existing challenges/gaps in IAF's ability to evaluate the programs it delivers?
    a. What is IAF hoping to achieve through more robust program evaluation practices?
    b. What elements would make for a comprehensive and replicable evaluation framework at IAF?
5. What data is IAF gathering to assess programs and is it the right data?

    a. What additional data and information should be collected to improve IAF's program evaluation, and why would these data/information improve program evaluation?
6. What is IAF's current capacity for collecting and analyzing the additional data? What is needed to improve IAF's capacity for collecting and analysing this data?
7. Is there anything else I should be asking?