

PROMISES AND PITFALLS:
NONPROFIT PROGRAM EVALUATION IN A COLLABORATIVE CONTEXT

A Thesis

Presented to the faculty of the Department of Public Policy and Administration
California State University, Sacramento

Submitted in partial satisfaction of
the requirements for the degree of

MASTER OF PUBLIC POLICY AND ADMINISTRATION

by

Alex McCreddin

SPRING
2020

© 2020

Alex McCreddin

ALL RIGHTS RESERVED

PROMISES AND PITFALLS:
NONPROFIT PROGRAM EVALUATION IN A COLLABORATIVE CONTEXT

A Thesis

by

Alex McCreddin

Approved by:

_____, Committee Chair
Sara McClellan, Ph.D.

_____, Second Reader
Andrea Venezia, Ph.D.

Date

Student: Alex McCreddin

I certify that this student has met the requirements for format contained in the University format manual, and this thesis is suitable for electronic submission to the library and credit is to be awarded for the thesis.

_____, Department Chair
Edward Lascher, Ph.D.

Date

Department of Public Policy and Administration

Abstract
of
PROMISES AND PITFALLS:
NONPROFIT PROGRAM EVALUATION IN A COLLABORATIVE CONTEXT
by
Alex McCreddin

Program evaluation is an important topic for nonprofit organizations since a range of nonprofit stakeholders expect to see organizations progress in achieving their mission and social purpose. Nonprofits experience a variety of challenges evaluating the impact of their programs, including lack of evaluation knowledge, resources, time, personnel, and evaluation training. The purpose of this study is to determine what challenges and barriers nonprofit organizations experience when evaluating their programs within a collaborative context. Specifically, this study aims to gather experiences from the growing number of nonprofit organizations that belong to nonprofit collectives and that are engaged in collaborative work and collective impact initiatives.

In this exploratory case study, I conducted in-depth interviews with nonprofit practitioners whose organizations belong to a Northern California nonprofit collective. The interviews provided rich and grounded detail on the array of challenges that nonprofits experience evaluating the performance of both individual and collaborative programs. These challenges include limitations in organizational capacity to conduct program evaluation, limitations in evaluation skill, funding barriers, and coordination challenges. The findings also indicate that all stakeholders, including nonprofit practitioners, funders, and collective leaders,

may benefit from engaging in dialogue to collaboratively reduce program evaluation challenges, barriers, redundancies, and inefficiencies.

_____, Committee Chair
Sara McClellan, Ph.D.

Date

ACKNOWLEDGEMENTS

I am so thankful for my family, friends, and loved ones who supported me during my time in the program. I cannot express enough gratitude for my wonderful wife Sheena for showing me unconditional love, supporting me, and for keeping me grounded throughout the program. I am also so grateful for my parents for encouraging me to seize my goals and for giving me the opportunity to participate in a Master's program.

To my fellow PPA colleagues, cohort members, and the GASPPA board, I am so thankful for the time we spent together in the program. I could not have asked for a more supportive, talented, and engaging cohort to enjoy this experience with. Thank you for the friendship and mutual support you provided along the way.

Finally, an enormous thank you to all faculty and staff in the Department of Public Policy and Administration. I am so appreciative of your dedication to student success and boundless enthusiasm for the program. I am eternally grateful for the support of my committee chair and thesis advisor Sara McClellan and for my second reader Andrea Venezia. Thank you Andrea for providing such thoughtful guidance on research design and methodology. Thank you Sara for being such a caring advisor, mentor, and sponsor throughout my time in the program. I could not have done this work without you.

TABLE OF CONTENTS

	Page
Acknowledgements	vii
List of Tables.....	x
List of Figures	xi
Chapter	
1. INTRODUCTION	1
Introduction to Program Evaluation.....	1
Collaborative Programming and Collective Impact	2
Primary Research Questions and Hypothesis	3
Study Design and Methodology	5
Early Study Limitations.....	6
2. LITERATURE REVIEW	7
Literature Review Introduction.....	7
Status of Current Program Evaluation Research.....	8
Theme 1: Program Evaluation Theory and Design.....	9
Theme 2: Program Evaluation Applications	13
Theme 3: Organizational Capacity for Program Evaluation	16
Summary of Literature Review.....	17
3. QUALITATIVE RESEARCH METHODS.....	19
Overview of Study Design and Methodology.....	19
Rationale for Conducting In-Depth Interviews	20
Selection of Subjects and Approach to Conducting Interviews	21

Ethical Research Considerations	23
Approaches to the Analysis and Early Limitations of Findings	23
4. IN-DEPTH INTERVIEW FINDINGS	26
Overview of Research Findings.....	26
Organizational Characteristics and Evaluation Practices	27
Application of Program Evaluation Results.....	30
Program Evaluation Challenges within Individual Organizations	32
Challenges Evaluating Collaborative Programs.....	35
Solutions and Desired Supports for Improving Program Evaluation Practices	37
Conclusion and Summary of Findings.....	39
5. DISCUSSION AND CONCLUSION.....	40
Analysis of Research Findings	40
Implications for Practitioners and Funders	43
Study Limitations and Opportunities for Future Research.....	46
Appendix A. Interview Protocol.....	47
References	49

LIST OF TABLES

Tables		Page
1.	Why do organizations evaluate their programs?.....	28
2.	What types of data do organizations collect?.....	29
3.	How do organization use and apply program evaluation results?.....	30
4.	How do organizations share program evaluation results?.....	31
5.	What challenges do organization face when evaluating their programs?.....	34
6.	What challenges do organizations face evaluating collaborative programs?.....	36

LIST OF FIGURES

Figures	Page
1. Main perspectives of nonprofits' performance.....	9

CHAPTER ONE: INTRODUCTION

Introduction to Program Evaluation

Program evaluation is an increasingly important topic for nonprofit practitioners, given the unique nature of nonprofit funding mechanisms and the diverse array of players with a stake in nonprofit program performance. Specifically, nonprofit organizations rely on charitable donations, grants, and government contracts as their primary sources of revenue. Therefore, donors and grantors, in addition to the communities served, are typically the primary stakeholders for nonprofit organizations. As Moore (2000) asserts, nonprofit organizations produce value by achieving their intended social purpose. Therefore, donors and other stakeholders expect to see nonprofits progress in achieving their mission, goals, and desired social outcomes. This expectation places particular importance on the role of program evaluation because it is the primary tool for organizations to measure their achievement of value.

Literature indicates that most nonprofits engage in at least one type of program evaluation activity, though currently most nonprofits continue to collect program output and activity data rather than actual social impact outcomes (Carman, 2007). However, recent case studies on perspectives of nonprofit leaders suggest a shift and desire for agencies to adopt more rigorous program evaluation practices that measure and improve program outcomes rather than just track outputs data (Alaimo, 2008). Despite this desire, research also indicates that many nonprofit organizations are limited in their capacity to conduct quality program evaluation due to lack of knowledge, resources, time, personnel, and training (Carman and Fredricks, 2010).

The impetus to measure program outcomes with high-quality data collection methods applies to both individual nonprofit organizations and the growing number of nonprofits that belong to collaborative partnerships. However, few studies examine how members of nonprofit collaborations conduct program evaluation. Even fewer studies discuss in detail the quality and

accuracy of program evaluation methods and whether organizations are collecting data that are both actionable and useful for invoking larger organizational or programmatic change. Therefore, my goal with this study is to learn what program evaluation challenges and opportunities exist within a collaborative context, with the goal of benefiting nonprofit practitioners and stakeholders, and academics studying nonprofit program evaluation.

Collaborative Programming and Collective Impact

Nonprofit collaboration has been a growing topic since the early 2000s as more organizations form alliances to address society's more complex and challenging problems (Acar & Guo, 2005). Specifically, collaboration is defined as multiple organizations coming together to search for and attempt to implement solutions to problems that would extend beyond their individual means. This collaborative work may entail conducting joint programs, pooling resources, engaging in shared decision-making, and sharing ownership of the program's impact. Studies have also attempted to categorize the different types of collaborative efforts. Acar and Guo assert that collaborations vary primarily based on the levels or degrees of engagement and collaboration intensity. Kania and Kramer (2011) describe collaboration typology further and suggest that collaborations come in five different forms: (1) funder collaboratives, (2) public-private partnerships, (3) multi-stakeholder initiatives, (4) social sector networks, and (5) collective impact initiatives.

Collective impact is important given that collaborative efforts attempt to *move the needle* on solving complex problems through collaboration. Kania and Kramer (2011) describe collective impact as a commitment among a group of important actors and organizations to solve specific social problems. As Emerson and Nabatchi (2015) describe, the demand for increased collaboration and collective impact is largely driven by the emergence of more complex *wicked problems* that are difficult to solve through individual efforts because of incomplete or

contradictory information, rapidly changing environments, and organizational interdependencies. These challenges or problems are commonly tied to health care, education, justice system, transportation, climate change, and poverty issues.

Identifying factors that contribute to collaboration success becomes important as more funders and organizations shift from supporting isolated impact to collective impact (Landers, Price, & Minyard, 2018). Landers et al. (2018) assert that factors such as establishing a common agenda, facilitating mutually reinforcing activities, fostering continuous communication among partners, providing backbone support, and establishing shared measurement systems are critical for successful collaborative initiatives. The concept of establishing shared measurement systems is particularly relevant for this study. Given that nonprofit organizations already experience challenges measuring impact on an individual basis, program evaluation becomes even more complex and interesting within a collaborative context. Thus, the purpose of this study is to learn what challenges organizations engaged in collective impact experience when evaluating program success.

Primary Research Questions and Hypothesis

This study focused on nonprofit practitioner perspectives on the challenges and barriers associated with program evaluation and data collection. My primary research question driving this study was *what practical challenges do nonprofit organizations encounter when they work to establish program evaluation practices for collaborative programs?* In addition to perspectives related to collaborative program evaluation, I also gathered perspectives on evaluation challenges related to individual organizational evaluation practices. Collecting experiences related to both individual and collaborative evaluation challenges was useful for comparing factors that overlap both individually and collectively, and factors that are unique specifically to collaborative

programs. In addition to my primary research question, I used several other sub-questions to guide this study:

- What is the primary impetus for organizations conducting program evaluation?
- What program evaluation designs and methodology do organizations employ?
- What sources of data do organizations collect for program evaluation?
- What challenges do organizations encounter when establishing indicators and metrics for evaluating program performance?
- How do organizations use and apply the results, if at all?
- What types of support do organizations need to conduct more effective or higher quality program evaluation?

My hypothesis was that nonprofit leaders would perceive a lack of resources, time, and technical knowledge as among the greatest barriers to conducting quality program evaluation. Additionally, I anticipated that these challenges would be more pronounced within the context of collaborative program evaluation. I based this hypothesis on several studies that suggest that resource, skill, and capacity challenges can limit an organization's ability to measure program performance. For example, Carman and Fredricks (2010) conducted a study on perceived challenges associated with program evaluation and the results indicate that the greatest challenges are lack of time (68%), lack of funding (51%), and lack of evaluation expertise (50%).

I also based this hypothesis on my own professional experience and personal anecdotes as a nonprofit program manager and youth program evaluator. As a practitioner, I have personally experienced evaluation challenges and limitations that coincide with and support Carman and Fredricks' (2010) findings regarding barriers to program evaluation. With over ten years of experience in the nonprofit sector, I have encountered a range of challenges associated with data collection, measuring program impact, and applying evaluation results to drive meaningful

programmatic or organizational change. Additionally, my experience interacting with a variety of nonprofit stakeholders has given me insight into the capacity and logistical challenges that can result from overlapping or cumbersome evaluation requirements for funded programs. I acknowledge that my professional experience related to program evaluation could introduce bias into the study. However, I worked collaboratively with Professor Sara McClellan during the coding process to ensure that analysis is accurate and as free of bias as possible. Thus, my own perspectives and experiences as a practitioner, and the perspectives described in the literature, provided important insight for this study.

Study Design and Methodology

I designed this study as an inductive qualitative case study with the purpose of gathering rich and grounded perspectives on evaluation challenges from nonprofit practitioners. The primary research site for this study was an emerging Northern California nonprofit collective that represents over 25 nonprofits primarily serving youth and families. To gather perspectives from partner agencies in the collective, I conducted 11 in-depth interviews with practitioners at ten different nonprofit organizations. During the in-depth interviews, I asked subjects open-ended questions to provoke rich and deep dialogue on perceived evaluation challenges both within an individual and collaborative program context. Prior to conducting the in-depth interviews, I also observed the collective's leadership meetings and partner activities for a one-year period. These experiences, though not formally incorporated into the study, provided me with the necessary background and understanding of the collective to guide my study design and interview protocol.

Professor McClellan and California State University, Sacramento helped facilitate my access to this research site since I was able to leverage their relationship with the collective's leadership. Prior to my involvement with this study, Professor McClellan and other students representing the university's Public Policy and Administration program had established and

maintained an ongoing relationship with the collective and assisted collective leadership with other administrative topics. Thus, these preexisting relationships helped me gain access to the research site.

Early Study Limitations

I anticipate challenges generalizing my findings given the design of this study. This is because it is typically challenging to generalize the findings from case studies and fieldwork (Singleton & Straights, 2010). The smaller sample size of participating subjects within one nonprofit collaboration may also increase challenges in generalizing the findings. However, the open-ended nature of this exploratory study yielded intriguing results on the perspectives of nonprofit leaders that other studies have not previously captured. This qualitative study also provided a greater context to explain *why* agencies belonging to nonprofit collaborations may experience challenges in evaluating programs.

CHAPTER TWO: LITERATURE REVIEW

Literature Review Introduction

Since the 1990s, stakeholders of nonprofit organizations have increased their calls for nonprofits to evaluate how effective their programs are at achieving their mission, purpose, and goals. Carman (2010) asserts that accounts of nonprofits mismanaging funds, using questionable fundraising practices, and engaging in fraud, embezzlement, and corruption have all led to stakeholders calling for increased accountability. The literature also suggests that the mechanisms nonprofits use to raise revenue, such as government and community grants and charitable contributions, provide the impetus for nonprofit accountability. This means that nonprofit leaders must be responsive to the expectations of funders and demonstrate that they are delivering value to the public by achieving their mission and social purpose (Moore, 2000).

Recent studies, however, suggest a shift to organizational learning and maximizing social impact as primary factors driving program evaluation activities within agencies. The literature also indicates that organizational capacity, including resources, personnel, time, and knowledge, may affect an agency's ability to design program evaluation activities and apply the results. This may be why more organizations track program outputs than social outcomes. However, as described above, recent literature indicates a shift to more organizations seeking to understand, measure, and improve the impact and outcomes of their work.

Lee and Nowell (2014) state that it is often challenging for nonprofit organizations to assess how successful they are in making an impact and achieving their mission and goals. These challenges in measuring programmatic success have led public officials, nonprofit leaders, and academics to call for increased research on nonprofit program evaluation. As Griffith and Montrosse-Moorhead (2017) describe, the field of program evaluation faces an "unprecedented, growing, global demand" (p. 1) with specific concerns for validity, reliability, and quality of

performance measurement efforts. Though program evaluation research has occurred since the late 1970s, the field has experienced the most growth since the 2000s. To demonstrate this, Coryn et al. (2017) summarize the state of program evaluation research from 2005 to 2014 and state that there have been 257 individual studies published across 14 journals focused on program evaluation.

Status of Current Program Evaluation Research

In the abundance of literature surrounding program evaluation, research mainly focuses on three themes, including program evaluation theory and design, how nonprofit organizations use and apply program evaluation results, and the organizational capacity to conduct program evaluations. In the following sections, I analyze program evaluation studies related to these three themes and discuss relevant findings, shortcomings in the research, and opportunities for future research. This analysis forms the basis of my research on program evaluation and provides a foundation for my question *what practical challenges do nonprofit organizations encounter when they work to establish program evaluation practices for collaborative programs?*

Overall, much of the research on program evaluation is in the form of case studies on individual nonprofit organizations, with an emphasis on their practices related to the above three themes. However, many studies do not provide specific details on the quality of program evaluation activities in regards to validity and reliability. Additionally, few studies on program evaluation include nonprofits belonging to collaborative partnerships or coalitions in their samples. There are even fewer studies on collaborative program evaluation efforts among nonprofit partners, leaving room for future research on these topics. I discuss these three themes in more detail below.

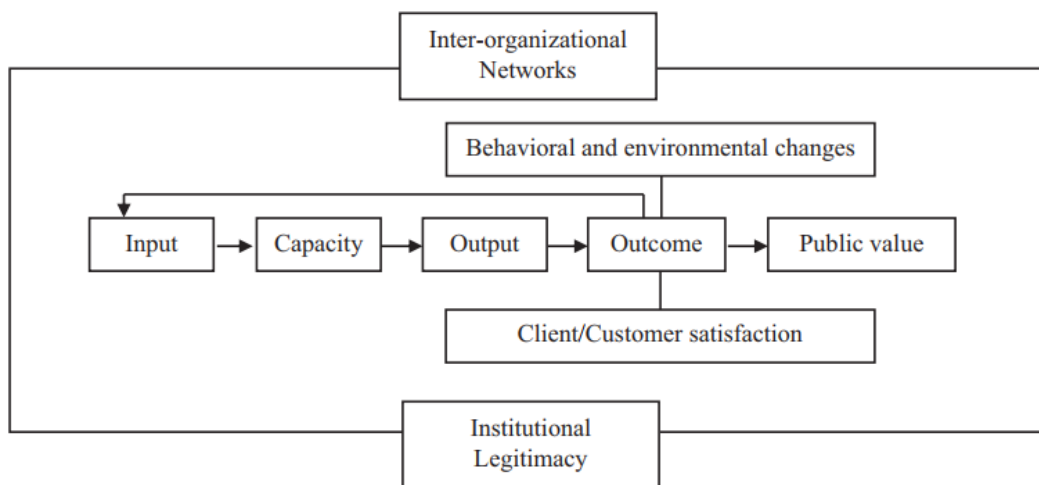
Theme 1: Program Evaluation Theory and Design

Program evaluation designs are rooted in theoretical models that describe how programs create social or environmental change. Thus, I begin this section by discussing several prominent theories surrounding programmatic change and the roles these theories play in shaping the program evaluation process. I then discuss the types of program evaluation activities agencies may engage in and the indicators that organizations may use to measure programmatic success.

Theory of Change and Programmatic Outcomes

Carman (2010) discusses the “Theory of Change” as a foundation for understanding how a program works and how its activities result in participant or environmental change. Evaluators or program leaders can articulate how a program works and produces results by creating a logic model of the program. In its most basic form, a logic model consists of the inputs or resources that a program uses, the activities of a program, and its outcomes related to the program’s social goals and purpose. Lee and Nowell (2014) provide a more comprehensive discussion on how programs produce change or outcomes, and depict this research with the below model.

Figure 1: Main perspectives of nonprofits’ performance (Lee and Nowell, 2014, p. 6)



Lee and Nowell's (2014) model provides additional emphasis on programmatic outcomes and ties that component to the public value that nonprofits create which leads to legitimacy with their stakeholders. Lee and Nowell's commentary provides a helpful way of distinguishing between outputs and activities, and program outcomes. They claim that outcomes differ from outputs because their reach is beyond activities and "seek to discern the impact of these activities on the target population" (Lee and Nowell, 2014, p. 8). The literature suggests that more organizations may currently track activities and outputs rather than outcomes, with one possible reason being challenges in establishing valid and reliable indicators to measure outcomes.

Types of Program Evaluation Activities

Two general questions typically drive the design of program evaluation methods, including "does a program work?" and "how does a program work?" (Constance & Solmeyer, 2015). In this section, I discuss program evaluation design aspects that primarily address the former question. By focusing on whether or not a program works, an agency leader or evaluator assesses how successful an organization is in achieving its mission and purpose.

Hoefler (2000) defines three types of program evaluation activities, including (a) implementation monitoring to check that the agency is delivering programs to the target population, (b) process evaluation to determine why a program did or did not achieve its outcomes, and (c) outcome evaluation to determine if an agency achieved its programmatic goals. Davis and Mizrahi (2008) expand on the types of evaluation activities that nonprofits may conduct with the additions of needs assessments of the target population and cost-efficiency evaluations of programs. The literature generally suggests that the majority of nonprofits engage in at least one of the above types of activities, however, implementation and quality of these activities may vary depending on the agency's accountability obligations and capacity to undertake program evaluation activities.

Indicators, Measures, and Data Sources

To evaluate the success of programs, nonprofits must establish performance indicators and incorporate a variety of data sources as evidence of meeting these indicators. Much of the literature surrounding program evaluation metrics suggests that indicators and measures should be both concrete and quantifiable. However, other studies suggest that a combination of both quantitative and qualitative evidence can provide greater context for evaluating a program's impact.

Greenwood (2008) describes indicators as “descriptors of something that is valued, they are all quantifiable, and they are all actually measureable” (p. 55). Greenwood also cautions against adopting indicators that are too broad of an aggregate to discern the program's impact. For example, teenage pregnancy rates, voter participation, or unemployment rates may all align with the goals of a program, but may be too high-level to assume causality from the program. As a general practice, Greenwood emphasizes that indicators should be as narrow and concrete as possible. However, in a discussion on objective versus value-based indicators, Brodsky (2014) asserts that over-relying on objective or concrete measures may result in the evaluation losing the larger context of the program, and advocates that evaluators use a combination of both types of indicators. Though Greenwood and Brodsky describe different approaches to establishing indicators, the commonality is that they should reflect the larger goals of the organization and should be specific enough to attribute client or environmental change to the program.

Regarding data and evidence, the literature includes both theoretical discussion and case studies on how nonprofits collect data. Buckmaster (1999) asserts that data collection methods should be both reliable and valid, which is a theme that occurs throughout the literature. Like Brodsky (2014), Buckmaster also advocates for both quantitative and qualitative evidence of program performance. Berriet-Sollic, Labarthe, and Laurent (2014) discuss different types of

data and conclude that evidence can either support the existence of a program or that the program made a change in the participant or environment. For example, data that supports the existence of a program may be attendance records or rosters and falls in the category of monitoring evaluation that Hoefler (2000) describes. Evidence and data capturing participant change is more challenging to quantify and warrants close attention to the validity and reliability of collection methods. This type of data is more often attributed to program outcomes rather than the program outputs that monitoring activities capture.

Case Studies on Program Evaluation Design and Activities

Several case studies aim to identify the program evaluation designs and types of data collection methods that organizations use most in practice. However, few studies include details on the quality, validity, and reliability of efforts, especially among studies of nonprofits that are members of collaborative coalitions. Hoefler (2000) surveys a sample of 91 human services agencies to learn what types of evaluation activities they engage in, and reports that 66% conduct evaluations related to monitoring program usage and 82% conduct outcomes-related evaluations. Additionally, Hoefler's study is one of the few that includes data on the quality and rigor of the program evaluation designs that agencies use. They find that some agencies incorporate a post-test-only design with weak internal validity (26%), a majority of agencies use a moderately rigorous pre- and post-test design (49%), and far less agencies incorporate a highly rigorous comparison group design with the highest internal validity (17%).

Botcheva, Huffman, and White (2002) gather data on a sample of 25 nonprofit organizations and is one of the few examples of studies that include members of nonprofit coalitions in their sample. Botcheva et al. find that the three most common data types that the nonprofits collect are client demographics (74%), client problem type (61%), and service outcomes (57%). Botcheva et al.'s study suggests that many organizations collect program

outcome data in addition to monitoring data. They find, however, that few agencies (22%) collect data related to long-term program outcomes. Though Botcheva et al.'s study includes members of nonprofit collaborations in their study, it is limited by a small sample size and lack of specificity regarding the quality, validity, and reliability of data collection methods.

Carman (2007) provides an expansive study on data collection in terms of both sample size and the scope. Carman surveys 178 nonprofits and finds that 84% of them conduct monitoring activities and fewer of them conduct activities related to program outcomes. For example, only 65% of agencies conduct formal program evaluations and only 41% use a performance measurement system. Among methods of collecting data, the three most common are written instruments, in-person interviews, and observations. Contradicting the results of Hoefler's (2000) study, Carman's findings indicate that most nonprofit organizations collect input and output data, while fewer nonprofits collect program outcomes data. In the organizational capacity section of this literature review, I discuss possible reasons for this gap in outcomes evaluation, such as technical knowledge, resources, personnel, and funding.

Theme 2: Program Evaluation Applications

Several concepts in the literature guide how an organization may apply program evaluation findings, including both evaluation policy and organizational learning tendencies. Christie and Dillman (2016) define evaluation policy as "any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation" (p. 2). Agencies may develop evaluation policies that inform several decisions related to the application of program evaluation results, including the audience of program evaluation activities, how agencies act upon results, and how agencies communicate results to stakeholders. Alternatively, as Christie and Lemire (2019) outline, the purpose of an agency's evaluation policy is to provide guidance on "how, when, in what way, and with what purpose evaluations are carried out" (p. 5). This

suggests that an agency's evaluation policies may play a role in shaping an agency's application of evaluation results, as these policies can help agencies establish and define the role that program evaluation plays in the organization.

A recurring theme in the literature about possible roles of program evaluation is organizational learning. For example, Botcheva et al. (2002) describe a learning organization as one that is open to change and continuous improvement. Alaimo (2008) expands on this description and defines organizational learning as "learning from success and failures, and changing behavior due to encountering situations within an organization's environment" (p. 77). Several recent studies indicate that organizational learning and continuous improvement could be a growing reason that many organizations conduct program evaluation. These recent findings may contradict previous theoretical discussion that describe stakeholder accountability as the primary impetus for conducting program evaluation that, for example, Moore (2000) and Carman (2010) discuss.

Case Studies on Program Evaluation Applications

Several case studies focus specifically on the impetus for conducting program evaluation, how agencies use evaluation results, and the audience of the results. These studies garner perspectives from the point of view of management personnel, executive leadership, and funders. However, the results share a commonality that most organizations conduct program evaluation activities to improve programmatic outcomes, and that many of these practices are informed by an organization's evaluation policy. Though the literature includes a variety of perspectives on the application of program evaluation results, there is a gap in the research due to a lack of studies on how members of nonprofit collaborations apply results.

Coghlan, Fine, and Thayer (2008) conduct an in-depth profile on four nonprofit organizations to discover why agencies conduct program evaluation, what influences stakeholder

involvement, and what contributes to the evaluation's usefulness. The results indicate that reporting to a funding agency (76%), improving outcomes or impact (75%), and changing program design or operations (73%) are the primary reasons for conducting program evaluation. The study also suggests that current funders (69%) and program staff (61%) are the primary audiences for the findings. Alaimo's (2008) study incorporates perspectives from 42 nonprofit executive directors and contradicts Coghlan et al.'s (2008) findings regarding the primary impetus for conducting program evaluations. The majority of executive directors that Alaimo interviews state that they use evaluation information to alter or improve their programs (67%). Executive directors also cite being responsive to stakeholders such as funders (17%) and staff (11%) as additional reasons for evaluating programs.

Christie and Dillman's (2016) study focuses on program evaluation applications from the perspective of a foundation that funds nonprofit evaluation efforts. The study points to several reasons for conducting program evaluation activities, including understanding the impact of programs, spreading effective approaches, ensuring credibility, and promoting social change. As is similar with other studies in the literature, Christie and Dillman's findings emphasize impact and organizational learning as a primary impetus for program evaluation. A subsequent study by Christie and Lemire (2019) also supports these findings, with two out of three nonprofits profiled emphasizing organizational learning as the primary impetus for program evaluation over accountability.

Overall, recent literature suggests a change in why nonprofit organizations evaluate programs and how the results are used. Earlier literature suggests stakeholder accountability, transparency, and legitimacy as the primary impetus for conducting program evaluation.

However, as studies from Alaimo (2008), Christie and Dillman (2016), and Christie and Lemire

(2019) indicate, the desire for organizational learning, continuous improvement, and increasing impact may be more significant than before for driving program evaluation.

Theme 3: Organizational Capacity for Program Evaluation

Overview of Evaluation Capacity Building (ECB)

Substantial literature emerged in the late 2000s regarding the Evaluation Capacity Building (ECB) efforts of nonprofit organizations. Wing (2004) helps distinguish between program evaluation and ECB, and states that the intended outcome of organizational ECB efforts are for internal agency change and improvements, rather than external client change. Blashki et al. (2007) define evaluation capacity as simply “the ability to conduct an effective evaluation” (p. 232), while Huffman, Lawrenz, and Thomas (2008) emphasize that ECB focuses on continuously working to develop organizational processes to make program evaluation a routine within an agency. Overall, the literature indicates that developing sustainable evaluation practices is a major goal of ECB and possible reasons for lack of success in these efforts may be limited agency capacity, resource limitations, and insufficient internal knowledge.

Many studies on ECB focus on what strategies agencies use to build sustainable evaluation practices internally. For example, Boyle and Preskill (2008) describe several mechanisms for ECB, including trainings, coaching, mentoring, technical assistance, written resources, and technology. In their model of organizational learning capacity, they state that program evaluation skills, knowledge, and internal attitude change regarding program evaluation must exist for an agency to develop sustainable evaluation practices. Huffman et al. (2008) describe similar practices for building organizational capacity to conduct evaluation, including technical assistance, workshops, and fellowships. However, of particular interest in their model is the Collaborative Immersion Approach to ECB, which indicates that collaborative ECB efforts may be more effective at building sustainable evaluation practices. These practices may include

collaborative data collection, data analysis, and action planning. Despite the many approaches to ECB, training and professional development to build sustainable evaluation practices appears to be a key theme in the literature.

Case Studies on Evaluation Capacity Building (ECB) Efforts

Several studies aim to discover how, in practice, agencies build internal program evaluation capacity and what barriers exist to developing these practices. Agius and Sobeck (2007) conducted a case study on 23 organizations and assert that the top three ECB activities that agencies participate in are mentorship, leadership circles, and workshops. The intended outcomes of these activities are to build managerial competencies, increase long-term program outcomes, and develop sustainable evaluation practices. Blashki et al. (2007) discuss the ECB efforts of an Australian agency, and state that written manuals, technical assistance, workshops, training of trainers, and using technology are the most common approaches to ECB.

In a later study, Carman and Fredricks (2010) describe not only how agencies conduct ECB but also what challenges and barriers agencies encounter when building evaluation capacity. Among the top three barriers to developing evaluation capacity are lack of time (68%), lack of funding (51%), and lack of evaluation expertise (50%). This study and other patterns in the literature suggest that most agencies have a desire to continuously develop, improve upon, and sustain effective program evaluation practices. However, possible barriers to building these practices could be lack of financial resources, personnel, time, technical skills, knowledge, technology, and internal leadership capacity.

Summary of Literature Review

Several patterns in the literature on program evaluation suggest that nonprofit organizations could benefit from additional research on program evaluation practices. It is evident that most nonprofit organizations practice at least one type of program evaluation activity.

However, studies indicate that more nonprofit organizations currently track program outputs and activities rather than social impact outcomes. Despite this tendency, case studies over the last 20 years indicate that more nonprofit leaders desire to track program outcomes to continuously improve the impact of their programs, rather than conducting program evaluation just for accountability purposes. However, an organization's knowledge, training, personnel, and resources may limit their capacity to develop high quality, valid, and reliable program evaluation practices.

Overall, gaps in the literature point to several opportunities for future study on program evaluation. First, few studies discuss the quality and rigor of program evaluation practices in regards to validity and reliability. Additionally, most case studies include individual nonprofit organizations in their sample rather than members of nonprofit collaborations. Given that a growing number of nonprofits belong to such coalitions, the field of program evaluation could benefit from future research discussing these practices in the context of nonprofit collaboration.

CHAPTER THREE: QUALITATIVE RESEARCH METHODS

Overview of Study Design and Methodology

Case studies indicate that organizations use a variety of approaches to evaluate their programs in practice (Hoefler, 2000; Botcheva, Huffman, & White, 2002; Carman, 2007).

Literature also suggests that these practices fall into several distinct categories, including (a) implementation monitoring to track program outputs, (b) target population needs assessment, (c) program outcomes evaluation, and (d) process evaluation to determine why or why not a program achieved its outcomes (Hoefler, 2000; Mizrahi, 2008). Often, the type of evaluation approach depends on the information and resources available to organizations. For example, case studies indicate that factors such as personnel resources, funding, and evaluation expertise influence organizational practices (Carman & Fredricks, 2010). Given the importance of both the type of evaluation approach selected and the factors at play that influence organizational practices, the methodology for this thesis focused on the perspectives of nonprofit employees related to how these factors may result in challenges or barriers to effective program evaluation.

As outlined in Chapter Two, there are many studies on the program evaluation practices of individual organizations, but few studies exist on the experiences of organizations evaluating programs within a collaborative context. Given that few studies examine these practices, I employed an inductive qualitative approach to learn about partners' perceptions of evaluation challenges within a nonprofit collective in Northern California. This chapter discusses the methodology I used to conduct this grounded case study.

I completed this research in partnership with Professor Sara McClellan of Sacramento State University, who initiated both the research site and Institutional Review Board (IRB) review for this study. Professor McClellan provided advisory support throughout this process and during the interviews and assisted with coordinating with collective leadership to schedule interviews.

This study on nonprofit program evaluation practices compliments Professor McClellan's and Sacramento State University's existing research and support on related topics at this site.

In this study, I conducted in-depth face-to-face interviews with 11 nonprofit employees from ten different organizations, with supportive note taking assistance from Professor McClellan. I selected the Northern California nonprofit collective as a research site because it has a rich diversity of nonprofit organizations in varying stages of development. Collective leaders invited all partners within the collective to participate in the study, resulting in 11 individuals responding. All subjects within this study had either a role in or specific knowledge of their organization's program evaluation practices. The objective of the interviews was to gather firsthand information on subjects' evaluation experiences and any perceived barriers or challenges to program evaluation, both in their individual organization and within the context of collaborative programming. Thomas (2006) describes this methodology as an inductive approach. This approach derives theoretical ideas or concepts from raw data, rather than beginning with a hypothesis or theory to drive the research. Thus, I employ an exploratory approach to learn of any themes or concepts that might emerge from the interview data.

Rationale for Conducting In-Depth Interviews

Different methodological approaches have their strengths and limitations. I chose in-depth interviews because they can provide rich and detailed perspectives from interview subjects. Learning about these perspectives can provide important insight for collective leaders and stakeholders to enhance partner capacities and the overall impact of the collective. Singleton and Straits (2010) assert that in-depth interviews are descriptive and explanatory in nature, and that they can provide an "insider's view of reality" (p. 356). Almeida, Faria, and Queirós (2017) emphasize that in-depth interviews have the potential to provide more rich and detailed insights

than other methodological approaches. This is because interviews allow the researcher to solicit more in-depth responses and to *probe* for more information, if necessary.

There are several drawbacks to utilizing in-depth interviews to collect data. First, interviews are time intensive to conduct and the data requires substantial coding (Singleton & Straights, 2010). These time constraints also require researchers to use smaller sample sizes than they would if they were employing other methods. Additionally, the analysis of interview data is highly dependent on the questions the interviewer decides to ask and the interviewer's interpretation of the responses. However, analyzing interview data in partnership with Professor McClellan helped reduce interpretation bias and provided a consistent analytical approach throughout the coding process. Despite the other drawbacks, in-depth interviews are a desirable approach for the purposes of this study. Almeida et al. (2017) assert that small sample sizes can still provide meaningful and relevant data because they allow researchers to explore topics in greater depth. In this case, the interviews provided a greater context for understanding subjects' perceived barriers to program evaluation. I gained this contextual understanding by asking broad open-ended questions and encouraging subjects to speak freely about the topic by building a rapport through personal interactions.

Selection of Subjects and Approach to Conducting Interviews

Singleton and Straits (2010) emphasize the importance of carefully selecting subjects for in-depth interviews. Unlike other methods, researchers use a non-random process to select a smaller number of subjects for case studies. Subjects can result from preexisting relationships or social encounters and can include individuals, organizations, or communities. I chose members of a Northern California nonprofit collective primarily serving youth and families because its partners have firsthand experience conducting collaborative programs. The collective consists of over 25 partner agencies that all vary in size and stage of development. This diversity allowed me

to capture a variety of experiences and perspectives on program evaluation within a collaborative context.

Singleton and Straits (2010) assert that “gatekeepers” (p. 374) can play an important role in allowing researchers to gain access to interview subjects. In this case, leaders within the collective helped facilitate the process by contacting collective partners to encourage them to participate in the study. Professor McClellan’s existing relationship with collective leaders also helped with gaining access to subjects. Overall, 11 participants from ten different agencies responded to the request and agreed to participate in the study. I conducted nine of these interviews in-person. Time and geographical constraints required me to conduct the other two interviews over the phone. However, this communication format did not result in any significant limitations in gathering subject perspectives. Each interview lasted approximately 45 to 75 minutes and took place over a period of one month.

The interviews followed a protocol that I divided into two major sections. The first section contained nine questions related to program evaluation practices and experiences pertaining to the subject’s individual organization. The second section of the interview protocol contained four questions related to the subject’s experience evaluating programs within a collaborative context. Though the interviews revealed other compelling avenues of potential research, I did not deviate from the interview protocol. However, Professor McClellan and I did employ probing questions to gather more in-depth details on the subject’s experiences and we asked clarifying questions if the subject’s responses were unclear. The protocol contained questions that were broad and open-ended to ensure that the protocol was neutral and non-leading. I include the full interview protocol in “Appendix A”.

Ethical Research Considerations

Several ethical considerations regarding informed consent and subject anonymity applied to this study. Singleton and Straits (2010) emphasize that research subjects should receive enough information to make an informed decision on whether or not they should participate in a study. At a minimum, subjects should understand that participating in the study is voluntary and that they are free to withdraw at any time. They should also receive a description of any potential risks of participating in the study. To allow participants to exercise informed consent, subjects received an IRB-approved informed consent form that described any potential risks from participating in the study.

Additionally, researchers should also guarantee the anonymity and confidentiality of subjects in a study (Singleton & Straits, 2010). This means that researchers should remove any names or identifying information from data as soon as possible. Additionally, any reports that emerge from the research should not include any information that can be attributed to the individuals or organizations in the study. I followed these guidelines when handling interview data and promised interview subjects that I would safeguard their anonymity and privacy when writing any reports from the data. I emphasized to participants that any findings from this study would not be attributable to themselves or their individual organizations.

Approaches to the Analysis and Early Limitations of Findings

Analyzing interview transcripts is time intensive and can be prone to error if the researcher does not code the data carefully (Singleton & Straits, 2010). I employed several steps to ensure that I coded the data accurately. First, Professor McClellan assisted me in each interview by transcribing the session. I also recorded each session electronically and consulted the audio recording if any portions of the transcripts were incomplete. Next, I pre-coded each transcript by applying a color scheme to ten different categories, including responses related to

program evaluation practices, perceived challenges to evaluation, and application of evaluation results. After I color-coded each transcript, I summarized the themes that resulted from each interview and synthesized this data with themes and significant findings resulting from multiple interviews. Professor McClellan and I then compared these themes collaboratively to ensure an overall accurate and consistent analysis.

Themes resulting from the interview data indicate that nonprofit organizations employ a variety of practices to evaluate their programs. These practices reflect several categories of program evaluation methodology described in the literature, including program implementation and output monitoring, target population needs assessment, and program outcomes evaluation. Additionally, subjects described a variety of challenges related to program evaluation, both within their individual organizations and within the context of collaborative programming. These challenges fell into several distinct categories, including evaluation capacity and skill limitations, participant-related barriers, and challenges related to funder behavior and expectations. I describe these themes in more depth in the next chapter.

I observed some methodological limitations during this process. First, several organizations in the study indicated that they had not yet participated in or evaluated collaborative programs within the Northern California nonprofit collective. In cases like this, I instead captured their perspectives on what challenges they think agencies in the collective might experience evaluating collaborative programs or pilot projects, or their previous experiences evaluating other collaborative programs. Additionally, only ten agencies of the approximately 25 partners in the collective participated in the study. Despite this relatively small sample size, the in-depth nature of the interviews allowed me to gather rich and grounded details of participant experiences. Additionally, a wide array of different sized organizations with different evaluation capacities participated in the study, resulting in a diversity of responses and perspectives.

Singleton and Straits (2010) observe that, in general, qualitative case studies have limitations related to the generalizability of findings. However, Flyvbjerg (2006) asserts that formal generalizability can in fact be overrated and that case studies contribute significantly to the “collective process of knowledge accumulation” (p. 227). Flyvbjerg emphasizes the importance of case study research and suggests that other conventional critiques of case studies may also be misleading. For example, qualitative case studies can produce context-dependent knowledge that facilitates expert learning on a subject. Case studies can also contain narrative aspects that describe real life complexities in a way that other methodologies may not capture. This suggests that this grounded research methodology has the potential to provide rich and descriptive detail on the perspectives of individuals and organizations evaluating programs within a collaborative context.

CHAPTER FOUR: IN-DEPTH INTERVIEW FINDINGS

Overview of Research Findings

In this chapter, I discuss findings from my in-depth interviews with nonprofit practitioners. My primary goal in conducting in-depth interviews was to gain rich and descriptive detail on the challenges that nonprofit organizations experience when evaluating their programs, specifically within a collaborative context. I discuss first the impetus for organizations evaluating their programs, the methodologies they employ, and the data sources they use. Next, I describe how agencies apply evaluation results and discuss any organizational changes that the results have driven. I conclude this chapter by summarizing themes related to perceived challenges and barriers to evaluating programs and highlight potential solutions and desired supports that interviewees identified.

I structure discussion on perceived evaluation challenges into two distinct categories: (a) challenges and barriers pertaining to interviewees' specific organizations, and (b) challenges resulting from interviewees' experiences evaluating collaborative programs. Subjects describe several major challenges specific to their individual organizations, including a lack of program evaluation capacity, participant-related barriers, and challenges resulting from funder behavior and expectations. Within a collaborative programming context, interviewees described challenges resulting from differing levels of evaluation skill and capacity among partners in the collective, differing program evaluation philosophies, and coordination challenges. Subjects also shared their perspectives on what they believe could overcome evaluation challenges within their organizations and the collective, which include solutions to increase evaluation capacity, additional funding for evaluation, and supports for increasing evaluation skill and methodology.

Organizational Characteristics and Evaluation Practices

I organize this section into several subcategories, including the impetus for organizations evaluating their programs, the instruments and methodologies that organizations use, the types of data that organizations collect, and their experiences participating in collaborative program evaluation. To ensure subject anonymity, I do not describe organizational characteristics related to each agency's mission, programs, individual size, or include any other identifying information that could be attributed back to the interviewee or their organization. However, as I discussed in the previous chapter, all 11 interview subjects belong to ten different organizations belonging to a Northern California nonprofit collective. These organizations primarily serve youth and families, and all vary in their mission, size, and stage of development.

Interviewees cite a variety of reasons why their organizations evaluate their programs, many of which are consistent with what the literature describes. A notable majority of interviewees (73%) indicated funder obligations as the primary impetus for evaluating their programs. This result was expected and is consistent with other case studies in the literature (Alaimo, 2008; Coghlan, Fine, & Thayer, 2008). Seven out of 11 respondents also indicated maximizing program impact and effectiveness as a primary reason for program evaluation. This finding is consistent with more recent studies on program evaluation and suggests that more organizations may be evaluating their programs to increase impact, foster organizational learning, and support continuous quality improvement (Christie & Dillman, 2016; Christie & Lemirel, 2019). Interviewees described other reasons for evaluating their programs, which include maintaining program continuity, better understanding their target population, identifying staff training needs, and incorporating participant feedback into program design. I provide a visual that lists these findings in the below table.

Table 1: Why do organizations evaluate their programs?

<i>Evaluation Purpose</i>	<i># Interviewees</i>	<i>Percentage</i>
To fulfill funder obligations	8	73%
Maximize program impact and effectiveness	7	64%
Maintain program continuity	3	27%
Gain better understanding of target population	3	27%
Identify staff training needs	2	18%
Incorporate participant feedback into program design	1	9%

Interviewees also indicated using a variety of methods, practices, and data sources for evaluating their programs. A slight majority, or six out of 11, respondents indicated that their organization employs a pre- and post-test design to measure impact. Several organizations indicated using other approaches, for example, collecting qualitative data, such as participant stories, perspectives, and journal entries, implementing program subject-matter assessments, and conducting in-person program observations.

Several themes emerge from the interviews that are consistent with data collection practices cited in program evaluation case study literature. A larger majority of interviewees (82%) indicated that their organizations collect output data, such as program attendance or units of service. A majority of organizations (73%) also collect target population demographic data, either directly from the participants or from secondary data. These findings compare to the smaller majority (55%) of subjects that indicated their organizations collect outcomes-related data to measure program impact. This is consistent with other case studies which suggests fewer organizations may actually be measuring program impact and outcomes (Botcheva, Huffman, &

White; 2002; Carman, 2007). The other types of data that organizations collect include participant satisfaction and aggregate program impact data.

Table 2: What types of data do organizations collect?

<i>Data Type</i>	<i># Interviewees</i>	<i>Percentage</i>
Program outputs for monitoring implementation	9	82%
Target population demographics	8	73%
Program outcomes to measure impact	6	55%
Participant satisfaction	2	18%
Aggregate program outcomes	1	9%

Finally, I asked interviewees to describe their experiences evaluating programs within a collaborative context. Though all interviewees are a member of the Northern California nonprofit collective, a majority (55%) indicated having no experience evaluating collaborative programs. Interviewees who did indicate evaluating collaborative programs shared that this occurred either within the context of the Northern California nonprofit collective or in other external collaborative efforts. There may be several factors that explain the lower levels of collaborative program evaluation among collective partners. Some interviewees indicated that differing organizational sizes and capacities among collective partners might be a barrier to evaluating collaborative programs. Several interviewees also indicated that competing priorities, such as program implementation, took precedence over program evaluation. Finally, other interviewees specified that they participate in the collective for reasons other than collaborating on program delivery. These reasons include access to resources, such as office space, personnel, and funding, or for philosophical reasons, such as belonging to a network or community.

Application of Program Evaluation Results

Interviewees described a variety of ways that their organizations use and apply program evaluation results, though few of these behaviors seem consistent across all organizations. Five out of 11 subjects indicated that their organizations apply program evaluation results to make program design or curriculum changes. Other applications of results include identifying participant needs, staffing and site location decisions, identifying staff training needs, and adjusting program schedules. As one subject described, “We like to draw on the information just to reflect on programming. What went well, what didn’t, and then what we can change. And then of course we use that for reporting to funders...and internal quality improvement” (Interviewee 5, personal communication, October 24, 2019). Other subjects described a participant-centered application of results, and indicated using results to identify additional participant needs or to determine if individual participants should continue with a particular program.

Table 3: How do organizations use and apply program evaluation results?

<i>Application</i>	<i># Interviewees</i>	<i>Percentage</i>
Program design or curriculum changes	5	46%
Identify participant needs	3	27%
Staffing and site location decisions	3	27%
Identify staff training needs	3	27%
Adjust program schedules	2	18%

Subjects also listed a variety of intended audiences for evaluation results and methods for sharing them. The intended audiences that interviewees described vary widely, and include funders, staff and volunteers, board of directors, partners and other stakeholders, and the general

public. A majority of interviewees (55%) indicated that they share results in an annual report format. Other methods of disseminating results include sharing them on agency websites, email distribution and newsletters, advertising, tours and events, social media, and qualitative storytelling.

Table 4: How do organizations share program evaluation results?

<i>Intended Audience</i>	<i># Interviewees</i>	<i>Percentage</i>
Funders	4	36%
Staff and volunteers	3	27%
Board of directors	3	27%
Partners and other stakeholders	3	27%
General public	2	18%
<i>Method of Sharing</i>	<i># Interviewees</i>	<i>Percentage</i>
Annual report	5	55%
Agency website	4	36%
Email distribution and newsletters	3	27%
Advertising	2	18%
Tours and events	1	9%
Social Media	1	9%
Qualitative storytelling	1	9%

The in-depth interviews demonstrated that organizations use a variety of approaches to apply program evaluation results and share them with external audiences. The results also indicated a surprising contrast in funder expectations versus internal aspirations for program evaluation. One interviewee described how their organization applies evaluation results in the context of satisfying funder desires versus increasing program impact:

Unfortunately, there's a lot of 'you collect information and report it for the funder', and as long as you have met the metrics that you promised you are doing, then that's success to them. For me, it's not about just checking the box, its ensuring that that person's quality of life is better after they have left or graduated from the program. (Interviewee 8, personal communication, October 31, 2019)

I discuss these tensions and other challenges associated with program evaluation in the next section.

Program Evaluation Challenges within Individual Organizations

Overall, interviewees described challenges that fall into two distinct categories. These categories include: (a) challenges related to program evaluation within their individual organizations, and (b) challenges evaluating programs within a collaborative context. Some experiences that participants described appear prominently in evaluation literature, and include challenges related to organizational experience, capacity, and funding. Other challenges included participant-related barriers, challenges with funder expectations, and differing evaluation capacities and philosophies within a collaborative program evaluation context. I describe below the challenges that interviewees indicated related to their individual organizations.

A majority of interviewees (64%) described limited staff time, or staff *wearing multiple hats*, as a barrier to effective program evaluation. Particularly, subjects described challenges related to balancing program evaluation with implementing or managing the programs. Other

interviewees indicated limited program evaluation planning and preparation time as a barrier. As one interviewee described, “We’re building the plane and flying it, and we’re in it... You don’t always have the luxury of planning” (Interviewee 3, personal communication, October 23, 2019). Subjects also described several challenges related to skill and knowledge of effective program evaluation practices. For example, four out of 11 interviewees indicated challenges related to “asking the right questions” on evaluation instruments and two interviewees cited concerns about the accuracy of self-reported participant responses. As an interviewee described, “I think that we’re always looking to make sure ‘is this the right thing we should be asking? Could we ask the question a better way? Is there another tool that’s more reliable?’” (Interviewee 5, personal communication, October 24, 2019). Finally, two subjects described capacity issues related to lack of technology to use for data collection and input.

Subjects suggested several barriers related to program participants not completing evaluation surveys and other assessments. For example, four interviewees cited challenges implementing pre- and post-surveys and assessments due to declining participant attendance during the evaluation cycle. Other participant-related challenges that interviewees described include participant unwillingness to respond to questions because of their content, participant privacy concerns, participant reading level, lack of participant time and availability, and survey fatigue.

Finally, interview subjects described several challenges pertaining to funder expectations. For example, four out of 11 interviewees indicated that funding amounts are a limiting factor. Specifically, they described how funding opportunities often do not fully cover all costs required to run a program, including the evaluation component. As one interviewee described, “We get enough to pay for what we do, bare minimum... But the way our funding structure is, it makes it very difficult to do more than the bare minimum” (Interviewee 4, personal communication,

October 23, 2019). Three respondents also described challenges related to differing funder expectations for program evaluation, often resulting in agencies accommodating funders by using several different instruments to evaluate the same program. Finally, two respondents indicated that shorter one to two-year evaluation timelines create barriers to effectively evaluating a program's full impact.

Table 5: What challenges do organizations face when evaluating their programs?

<i>Skill and Capacity Challenge</i>	<i># Interviewees</i>	<i>Percentage</i>
Staff time limitations	7	64%
Instrument-related challenges	6	55%
Lack of technology for data collection and input	2	18%
<i>Participant Challenge</i>	<i># Interviewees</i>	<i>Percentage</i>
Declining attendance during evaluation cycle	4	36%
Unwillingness to answer questions	4	36%
Privacy concerns	3	27%
Reading level challenges	2	18%
Lack of participant time and availability	2	18%
Survey fatigue	2	18%
<i>Funder Challenge</i>	<i># Interviewees</i>	<i>Percentage</i>
Funding amounts do not cover program costs	4	36%
Differing funder expectations for instruments	3	27%

Short evaluation timelines	2	18%
----------------------------	---	-----

Challenges Evaluating Collaborative Programs

In this section, I outline the challenges that interviewees cited evaluating programs within a collaborative context. Several of these challenges are similar to the barriers that interviewees described evaluating their individual programs, particularly surrounding capacity and skill. Other challenges are specific to collaborative programming and include coordination issues and tensions resulting from differing program evaluation practices and philosophies among partner agencies.

A primary barrier that interviewees described is limited organizational time and capacity. Four respondents indicated that implementing the collaborative programming required most of their time and attention, which left little time for program evaluation. In describing their experience evaluating a collaborative pilot program, one interviewee described, “Evaluation was a challenge and almost non-existent because there was a lot of variables that we experienced for the first time, so it was kind of growing pains for us” (Interviewee 5, personal communication, October 24, 2019).

Three interviewees also suggested tensions and challenges evaluating collaborative programs due to differing agency skill levels and capacities. Surprisingly, interviewees representing both smaller and larger organizations described these tensions. An interviewee representing a larger and more established organization indicated:

From a national brand standpoint, I feel like I have to be really cautious about entering into anything new. And because we’re really known for our data and our integrity in that way, I have some fear, and I think our organization has some fear, about diluting some of that. (Interviewee 2, personal communication, October 11, 2019)

Alternatively, an interviewee representing a smaller organization described:

I don't want to be put in this position where we have to try to justify being in the same room with all these [larger] nonprofits that already have this capacity to be able to come up with data. That already have these systems in place. (Interviewee 11, personal communication, November 1, 2019)

Interviewees also discussed evaluation challenges related to partner participation, follow-through, and coordination. Four interviewees described tensions and challenges evaluating programs within the nonprofit collective due to member agencies holding differing philosophies on program evaluation. For example, one interviewee indicated less established agencies might be focused more on implementing the programs than evaluating them. Another subject suggested that organizations might not have an incentive to evaluate collaborative programming unless the evaluation activity is tied to funding. Three interviewees indicated challenges working collaboratively with other partners to evaluate programs because of lack of follow-through and data sharing. Finally, three subjects also described experiences with partners reporting or inputting data in inconsistent ways.

Table 6: What challenges do organizations face evaluating collaborative programs?

<i>Collaboration Challenge</i>	<i># Interviewees</i>	<i>Percentage</i>
Limited time and capacity for evaluation	4	36%
Differing evaluation philosophies among partners	4	36%
Tensions from differing organizational skills & capacities	3	27%
Partner follow-through and data sharing	3	27%
Inconsistent data reporting or input	3	27%

Solutions and Desired Supports for Improving Program Evaluation Practices

Overall, respondents suggested a wide variety of potential solutions to mitigate the challenges and barriers associated with individual and collaborative program evaluation. These solutions largely surround increased personnel and staffing support, additional funding for evaluation, support with improving tools and evaluation methods, and enhanced technology for data collection and input.

Three themes emerged when I asked interviewees what supports would be beneficial for improving evaluation methodology within their individual organizations. These supports include increasing general organizational capacity, improving evaluation skill and methodology, and additional funding for evaluation. Five interviewees indicated a desire for a dedicated program evaluation staff to both coordinate and implement evaluation efforts. As one interviewee described, “A lot of it is just human capacity...In a perfect world, I would have somebody who could just conduct [participant] assessments. That would be their job. They’re embedded in the program” (Interviewee 2, personal communication, October 11, 2019). Three interviewees also suggested that an external program evaluator could increase the accuracy of evaluation results. Another three interviewees also described a desire for increased funding, specifically to expand evaluation capacity or to improve instruments and methodology. Finally, two interviewees suggested that improving technology could make data collection and input more efficient. For example, one participant asserted that adopting technology to electronically input data would be desirable because inputting paper surveys can be cumbersome and can limit efficiency.

Within a collaborative context, interviewees indicated several solutions for increasing program evaluation capacity among collective partners. As is consistent with supports desired for their individual agencies, interviewees suggested increasing evaluation personnel as a solution for improving the collective’s program evaluation capacity. A majority (55%) of interviewees

suggested that a dedicated program evaluation lead or coordinator would be ideal for enhancing the collective's evaluation efforts. As an interviewee indicated:

You almost need somebody with a higher global view and who can help connect the pieces. I think having a person who is skilled in data and evaluation, and also has time to do it and could be paid to do that work, would be really helpful. (Interviewee 2, personal communication, October 11, 2019)

Interviewees suggested other personnel solutions, including creating a program evaluation committee within the collective and hiring a program evaluation trainer to enhance the evaluation capacities of partners.

Interviewees also described a variety of ideas for enhancing evaluation skill and knowledge among collective partners. Three respondents asserted that the collective and its partner agencies should develop a stronger vision regarding the importance of data. As one subject suggested, "I think it's the education piece. I think its reaching those nonprofit leaders, or whoever is administering programs, and really educating them about the power of data" (Interviewee 1, personal communication, October 10, 2019). Three interviewees indicated that increasing the collective's storytelling capacity is important for demonstrating the impact the collective is having on the community, "It's about storytelling. It's about deciding what story you want to tell with that data...The story that we tell is just as important as having the data" (Interviewee 9, personal communication, October 31, 2019). To compliment this suggestion, two interviewees also indicated that it is important for the collective to enhance its evaluation methods so it can demonstrate its aggregate impact on the community, or how it is *moving the needle* in regards to impact on the target population.

Conclusion and Summary of Findings

The in-depth interviews revealed a variety of factors compelling organizations to evaluate their programs, and differing aspirations and practices for applying the evaluation results. Primarily, these organizations appear to evaluate their programs for the purposes of satisfying funder requirements, for increasing the impact and quality of their programs, or for a combination of both satisfying funder requirements and increasing program quality. Interview subjects also expressed a variety of challenges and barriers related to effective program evaluation. The most notable consistencies in evaluation challenges include limitations in organizational capacity and staffing, challenges with methodology and evaluation design, and difficulties resulting from funder behaviors and expectations. These factors may exacerbate evaluation challenges within a collaborative programming context, leading to tensions related to differing partner capacities, skill levels, and evaluation philosophies, and general coordination challenges with data collection and input.

In the final chapter, I analyze these findings more closely and identify possible avenues for future research on nonprofit program evaluation. I also discuss potential strategies for reducing identified barriers and challenges associated with program evaluation. In the next chapter, I also emphasize how individual nonprofit organizations, collectives, and funders may all mutually benefit from identifying shared goals surrounding program impact and evaluation, aligning efforts to enhance organizational capacity and evaluation methods, and fostering mutual support and understanding among both organizations and funders.

CHAPTER FIVE: DISCUSSION AND CONCLUSION

The in-depth interviews provided rich and grounded detail on evaluation experiences and perspectives from 11 nonprofit practitioners whose organizations participate in a Northern California nonprofit collective. The findings indicated that nonprofits evaluate their programs for a variety of reasons, such as fulfilling funder requirements, improving program outcomes, or for continuous quality improvement. They also experience a range of evaluation challenges both within their individual organizations and in the context of collective programming. These challenges fall into several distinct categories, including limitations in organizational capacity to conduct program evaluation, limitations in evaluation skill, funding barriers, and coordination challenges. Therefore, the purpose of this concluding chapter is to analyze these findings and compare them to themes in evaluation literature. Additionally, I raise several questions to provoke continued dialogue surrounding the types of challenges and opportunities organizations face when engaging in and evaluating collective work, and raise several issues that nonprofit practitioners, funders, and stakeholders might benefit from addressing collaboratively.

In the first section of this chapter, I compare this study's findings with themes in evaluation literature. I also describe how these findings relate to my initial hypothesis described in the introductory chapter. In the next section, I discuss the implications this study may have for both nonprofit practitioners and funders. I conclude in the final section by suggesting opportunities for future research and note limitations I observed while conducting the study.

Analysis of Research Findings

This study generally yielded findings that are consistent with what other evaluation studies in the literature describe. I compare this study's findings to other evaluation case studies in terms of factors that drive nonprofit program evaluation, evaluation methodology and design, application of evaluation results, and challenges with evaluation implementation and data

collection. Despite thematic similarities, the study also yielded some surprising and unexpected results. Given the small and localized sample size of this study, some of these surprises may be the result of the unique nature of the research site. For example, the large population of nonprofits in the region that this research site is located in often results in organizations competing for scarce resources, which may exacerbate the evaluation challenges I note in this study. However, there are still insights that have relevance for future case studies on collaborative program evaluation that are applicable to both nonprofit evaluation practitioners and funders.

As expected, a large majority of interviewees indicated funder expectations and requirements are a primary impetus for evaluation practices. Despite this tendency, a central theme across a majority of interviews was that many organizations aspire to use evaluation as a tool for enhancing program performance and increasing the impact their programs have on the populations they serve. This finding may be surprising within the context of earlier literature. For example, Moore (2000) observes that legitimacy, transparency, and accountability are key themes driving discussion surrounding nonprofit performance and evaluation. However, the language that interviewees used to describe their goals surrounding program evaluation was consistent with an emerging evaluation literature which often highlights organizational learning, continuous improvement, and increasing impact as an impetus to evaluate (Alaimo, 2008; Christie and Dillman, 2016; Christine and Lemire, 2019).

Though a majority of organizations expressed a desire to use evaluation as a tool for continuous improvement and learning, I observed some inconsistencies in the efficacy of implementing these aspirations. For example, a larger majority (82%) of interviewees indicated that their organization collects program outputs and monitoring data compared to the 55% of interviewees that indicated collecting program outcomes and impact-related data. Even fewer interviewees described any significant organizational or programmatic changes resulting from

evaluation results. I anticipated these findings, however, because many evaluation case studies find that nonprofits more frequently track program outputs rather than using indicators to measure program outcomes or performance (Botcheva et al., 2002; Carman, 2007).

These implementation challenges fit into a larger pattern of challenges that nonprofit organizations experience when evaluating their programs. My hypothesis was that interview subjects would share evaluation challenges and barriers resulting from limitations in resources, time, and evaluation expertise. I also assumed that interviewees would describe these challenges with greater emphasis on the context of collaborative programming and evaluation. As anticipated, a majority of subjects described evaluation challenges arising from limitations in personnel time, other competing priorities that limit evaluation implementation, and a number of barriers related to expertise in methodology and instrument design.

However, I observed several unanticipated findings that link these challenges to funder behavior and expectations. As four interviewees noted, funding opportunities often do not fully cover all costs associated with running a program, including the evaluation component. Additionally, subjects noted that funding timelines are often too short to effectively measure the full impact of a program. These funding limitations may lead organizations to seek multiple funding opportunities to maintain a program's operations. As I described in the previous chapter, this tendency often results in organizations evaluating their programs with multiple tools to satisfy different funder requirements. This could lead to other participant-related barriers such as *survey fatigue* or time limitations for program participants to complete the evaluation instrument.

I also anticipated many of the collaborative evaluation challenges that interviewees described. For example, a majority of interviewees indicated similar challenges related to personnel, time, and overall capacity limitations in a collaborative context. Several interviewees noted that the *pilot project* nature of the collective's programs resulted in partners giving more

attention to program implementation rather than evaluation. This challenge of competing priorities may also be exacerbated for smaller organizations that are still developing their overall capacities. To overcome the complexities associated with collaborative program evaluation, interviewees commonly suggested that the collective embed a data specialist to coordinate evaluation activities and to provide training to enhance the evaluation expertise of its partners. This suggestion provokes a larger question as to how nonprofit collectives can build the evaluation capacities of their members and to what extent funding opportunities should cover costs associated with evaluation activities. It also mirrors discussion in other case studies on evaluation, particularly surrounding Evaluation Capacity Building efforts (Boyle and Preskill, 2008) and indicates that collectives may benefit from coordinating Evaluation Capacity Building efforts for partners.

Implications for Practitioners and Funders

This study identified several key issues surrounding nonprofit program evaluation and community-based collective programming efforts. I assert that many of these issues share commonalities in both an individual and collaborative evaluation context. Scarce or incomplete evaluation capacity, skill, and knowledge may limit effective program evaluation; however, many organizations desire to enhance their evaluation efforts to increase their impact on and understanding of the populations they serve. Nonprofit collectives are also tasked with solving society's more complex and *wicked* problems, which brings its own set of unique challenges given the diversity of skills, backgrounds, and understandings that different collective partners bring.

Additionally, this study revealed the complexities that come with implementing and measuring the impact of collective efforts. The findings demonstrate that nonprofit collectives may be challenged to meet the needs of smaller partner agencies that are still developing their

own systems, competencies, and culture, while still offering an incentive for larger and more developed partner organizations to participate. Additionally, a collective may not only play a role in developing the individual capacities of its partners, but also in developing its own universal indicators of impact and performance to demonstrate how it is *moving the needle* in impacting the larger community. Funders of nonprofit organizations and collectives also play a role in establishing program priorities and expectations and allocating resources for funded programs. This study suggests that there may be opportunities for all of these key players to engage and align values and priorities and develop strategies for fostering mutual support and understanding.

Stakeholders of community programming efforts who seek to enhance evaluation expertise might turn to Evaluation Capacity Building literature as a starting point. As described in the literature review chapter, the purpose of Evaluation Capacity Building is to build an organization's internal capacities to make program evaluation a routine and sustainable activity (Huffman et al., 2008). These types of activities typically include facilitating trainings, coaching, mentoring, offering technical assistance, providing written resources, and providing technology to enhance an organization's evaluation ability (Boyle and Preskill, 2008). Since nonprofit collectives themselves are a venue for community efforts, collectives may benefit from using this advantage to provide greater supports for their partners to develop sustainable evaluation practices. This strategy could also serve as a way to leverage the unique skills and abilities of all collective partners to create shared learning and collaborative Evaluation Capacity Building opportunities, and develop a shared vision for creating, measuring, and describing impact within the community.

Additionally, nonprofit collectives could also serve as a venue for convening all players involved in collaborative processes, including nonprofit and collective leaders, funders, and other stakeholders, to address these challenges. Convening key stakeholders in collaborative processes

and dialogue may provide an opportunity to discuss the tensions and opportunities that this study identified. For example, opportunities may exist for stakeholders to:

- Develop strategies to align funding opportunities to meet the needs of both funders and funded agencies related to evaluation methods, evaluation timelines, funding amounts, and desired impact.
- Discuss how to streamline evaluation processes to reduce redundancies and develop universal tools to make evaluation more efficient.
- Identify resources or funding to reduce personnel and capacity-related barriers to effective program evaluation.
- Identify Evaluation Capacity Building options to enhance the evaluation expertise of collective partners, such as providing trainings, workshops, coaching, or technical assistance.

Convening stakeholders involved in collaborative efforts could also open up opportunities for broader dialogue on the intent and purpose of program evaluation. As Moore (2000) and Lee and Nowell (2014) suggest, organizations create public value by achieving their ultimate social purpose. This suggests that both organizations and funders may benefit from collaborating to understand which evaluation activities are most important and useful for measuring achievement of that social purpose. By aligning intent behind evaluation activities, stakeholders may be able to reduce some of the challenges interviewees expressed in this study, such as redundant or logistically cumbersome evaluation requirements, or difficulties in applying evaluation results to invoke meaningful organizational or programmatic change. Thus, nonprofit practitioners and funders could both benefit from discussing what evaluation activities may be the most effective for producing results that are both relevant and actionable.

Study Limitations and Opportunities for Future Research

This exploratory study examined a diverse array of experiences related to nonprofit program evaluation. However, due to factors related to the design and scale of this study, I would caution against making larger conclusions or generalizations from the research findings, given the small sample size of participating subjects and agencies. As described previously, certain challenges and tensions described in this study may also be unique to the realities facing this specific region and community of nonprofit organizations and funders. However, the purpose of this narrative is to bring attention to the realities that nonprofit organizations and collectives may experience when working to solve some of society's more complex and *wicked* problems. It is also my hope that stakeholders engaged in collective efforts dialogue about these evaluation issues in an attempt to increase mutual understanding, support, and growth, and to develop strategies to enhance the impact of collective efforts.

Given the narrative detail and context that this study provided, I observed several opportunities for future research on nonprofit program evaluation and measuring collective impact. I hope this preliminary study incites more in-depth research on nonprofit collectives and their attempts to operationalize, evaluate, and increase program impact. Particularly, future researchers may consider studying these issues within emerging nonprofit collectives and conducting in-depth case histories on more established collectives. Additionally, researchers might consider exploring Evaluation Capacity Building efforts and opportunities within the context of community-based collective efforts, in an attempt to understand how nonprofits and nonprofit collectives can develop evaluation practices that are sustainable and provide relevant and actionable results to organizations. Finally, my hope is that this study expands the discourse on the roles that all stakeholders play within larger collective efforts, including organizations, collectives, collective leaders, funders, and research or evaluation partners.

APPENDIX A: INTERVIEW PROTOCOL

The questions in Section I pertain to your own nonprofit organization rather than the collective.

Section I:

- Please tell us how many employees are typically involved in conducting evaluation activities in your organization.
 - How many, if any, external contractors support your evaluation activities?
- Please describe the primary reasons your organization evaluates its programs.
- Describe the types of evaluation activities your organization uses to assess its programs.
 - **Follow-up clarification:** For example, needs assessment, posttest, pre/posttest, or longitudinal.
 - How often does your organization conduct program evaluations?
- Please describe how your organization establishes indicators or measures of program performance, if at all.
 - Does your organization encounter challenges when working to establish indicators and metrics for evaluating programs? If so, please describe them.
- What sources of data does your organization use to evaluate its programs?
 - What types of data does your organization collect? **Follow-up clarification:** For example, demographic, target population needs, client participation, or outcomes.
 - Who in your organization is responsible for collecting this data?
 - How does your organization collect the data? **Follow-up clarification:** For example, surveys, interviews, observations, or focus groups.
 - Does your organization encounter any challenges when collecting data? If so, please describe them.

- Please describe how your nonprofit organization uses and applies program evaluation results, if at all.
- Please describe how program evaluation results have driven any changes for your nonprofit organization.
- Describe how your organization shares program evaluation results, if at all.
 - **Follow-Up Clarification:** Who is the primary audience for the results?
- Please describe the type of support you think your nonprofit organization needs to effectively evaluate programs.

Section II:

The following questions in Section II pertain to the Youth & Family Collective rather than your individual nonprofit organization.

- Please describe any experience your organization has had in partnering with other nonprofits to establish shared evaluation activities.
- If your organization has partnered to conduct a collaborative pilot project with the Youth & Family Collective, please describe your shared evaluation activities.
 - What types of performance measures did you use?
 - How did you collect evaluation data?
 - What, if any, challenges have you encountered in conducting evaluation activities with collaborative partners?
- Please describe the type of support you think the collective will need to coordinate program evaluation efforts across the work of different nonprofit partners.
- Is there anything else you would like to share with us about your experience with collaborating on evaluation activities?

REFERENCES

- Acar, M., & Guo, C. (2005). Understanding collaboration among nonprofit organizations: Combining resource dependency, institutional, and network perspectives. *Nonprofit and Voluntary Sector Quarterly*, 34(3), 340–361. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/0899764005275411>
- Agius, E., & Sobek, J. (2007). Organizational capacity building: Addressing a research and practice gap. *Evaluation and Program Planning*, 30(3), 237-246. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0149718907000390?via=ihub>.
- Alaimo, S. P. (2008). Nonprofits and evaluation: Managing expectations from the leaders perspective. *New Directions for Evaluation*, 2008(119), 73-92. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.269>.
- Almeida, F., Faria, D., & Queirós, A. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*, 3(9). Retrieved from https://www.researchgate.net/publication/319852576_Strengths_and_Limitations_of_Qualitative_and_Quantitative_Research_Methods
- Berriet-Solliec, M., Labarthe, P., & Laurent, C. (2014). Goals of evaluation and types of evidence. *Evaluation*, 20(2), 195-213. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1356389014529836?journalCode=evia>.
- Blashki, G., Burgess, P., Kohn, F., Morley, B., Naccarella, L., & Pirkis, J. (2007). Building evaluation capacity: Definitional and practical implications from an Australian case study. *Evaluation and Program Planning*, 30(3), 231-236. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S014971890700033X?via=ihub>.

- Botcheva, L., Huffman, L. C., & White, C. R. (2002). Learning culture and outcomes measurement practices in community agencies. *American Journal of Evaluation*, 23(4), 421-434. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/109821400202300404>.
- Boyle, S., & Preskill, H. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459. Retrieved from <https://journals.sagepub.com/doi/10.1177/1098214008324182>.
- Brodsky, R. L. (2014). Commentary: "Public value" and the measurement of government performance: The shift to subjective metrics. *Public Administration Review*, 74(4), 478-479. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/puar.12234>.
- Buckmaster, N. (1999). Associations between outcome measurement, accountability and learning for non-profit organisations. *International Journal of Public Sector Management*, 12(2), 186-197. Retrieved from <https://www.emerald.com/insight/content/doi/10.1108/09513559910263499/full/html>.
- Carman, J. G. (2007). Evaluation practice among community-based organizations. *American Journal of Evaluation*, 28(1), 60-75. Retrieved from <https://journals.sagepub.com/doi/10.1177/1098214006296245>.
- Carman, J. G. (2010). The accountability movement: What's wrong with this theory of change? *Nonprofit and Voluntary Sector Quarterly*, 39(2), 256-274. Retrieved from <https://journals.sagepub.com/doi/10.1177/0899764008330622>.
- Carman, J. G., & Fredericks, K. A. (2010). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation*, 31(1), 84-104. Retrieved from <https://journals.sagepub.com/doi/10.1177/1098214009352361>.

- Christie, C. A., & Dillman, L. M. (2016). Evaluation policy in a nonprofit foundation: A case study exploration of the Robert Wood Johnson Foundation. *American Journal of Evaluation, 38*(1), 60-79. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1098214016642864>.
- Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation*. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1098214018824045?journalCode=ajec>.
- Coglan, A., Fine, A. H., & Thayer, C. E. (2000). Program evaluation practice in the nonprofit sector. *Nonprofit Management and Leadership, 10*(3), 331-339. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/nml.10309>.
- Constance, N., & Solmeyer, A. R. (2015). Unpacking the “black box” of social programs and policies: Introduction. *American Journal of Evaluation, 36*(4), 470-474. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1098214015600786>.
- Coryn, C. L., Fiekowsky, E. L., Greenman, G. D., Hobson, K. A., Ozeki, S., Schröter, D. C., Westine, C. D., Wilson, L. N. (2017). A decade of research on evaluation: A systemic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation, 38*(3), 329-347. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1098214016688556?journalCode=ajec>.
- Davis, L. E., & Mizrahi, T. (2008). Program evaluation. *Encyclopedia of Social Work*. Retrieved from <https://www.oxfordreference.com/view/10.1093/acref/9780195306613.001.0001/acref-9780195306613>.
- Emerson, K., & Nabatchi, T. (2015). Collaborative governance regimes. Washington, D.C.: Georgetown University Press.

- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. Retrieved from <https://journals.sagepub.com/doi/10.1177/1077800405284363>
- Greenwood, T. (2008). Bridging the divide between community indicators and government performance measurement. *National Civic Review*, 97(1), 55-59. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/ncr.207>.
- Griffith, J. C., & Montrosse-Moorhead, B. (2017). Toward the development of reporting standards for evaluations. *American Journal of Evaluation*, 38(4), 577-602. Retrieved from <https://journals.sagepub.com/doi/full/10.1177/1098214017699275>.
- Hoefler, R. (2000). Accountability in action?: Program evaluation in nonprofit human service agencies. *Nonprofit Management and Leadership*, 11(2), 167-177. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/nml.11203>.
- Huffman, D., Lawrenz, F., & Thomas, K. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation*, 29(3), 358-368. Retrieved from <https://journals.sagepub.com/doi/10.1177/1098214008318798>.
- Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*. Retrieved from https://ssir.org/images/articles/2011_WI_Feature_Kania.pdf
- Landers, G., Minyard, K., & Price, K. (2018). Developmental evaluation of a collective impact initiative: Insights for foundation. *The Foundation Review*. Retrieved from <https://scholarworks.gvsu.edu/tfr/vol10/iss2/10/>
- Lee, C., & Nowell, B. (2014). A framework for assessing the performance of nonprofit organizations. *American Journal of Evaluation*, 36(3), 299-319. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1098214014545828>.

- Moore, M. H. (2000). Managing for value: Organizational strategy in for-profit, nonprofit, and governmental organizations. *Nonprofit and Voluntary Sector Quarterly*, 29(1_suppl), 183-204. Retrieved from <https://journals.sagepub.com/doi/10.1177/0899764000291S009>.
- Singleton, R., & Straits, B. C. (2010). *Approaches to Social Research*. New York, NY: Oxford University Press.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1098214005283748?journalCode=ajec>
- Wing, K. T. (2004). Assessing the effectiveness of capacity-Building initiatives: Seven issues for the field. *Nonprofit and Voluntary Sector Quarterly*, 33(1), 153-160. Retrieved from <https://journals.sagepub.com/doi/10.1177/0899764003261518>.