

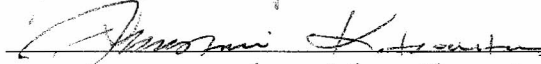
**NONPROFIT ORGANIZATIONS' EXPERIENCES WITH EVALUATION**

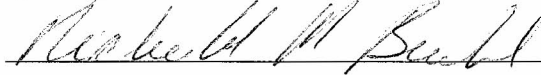
by

Ann K. Emery  
A Thesis  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Master of Science  
Educational Psychology

Committee:

  
\_\_\_\_\_ Chair

  
\_\_\_\_\_

  
\_\_\_\_\_

  
\_\_\_\_\_ Program Coordinator

  
\_\_\_\_\_ Dean, College of Education  
and Human Development

Date: May 2, 2014 \_\_\_\_\_  
Spring Semester 2014  
George Mason University  
Fairfax, VA

Nonprofit Organizations' Experiences with Evaluation

A thesis submitted in partial fulfillment of the requirements of the degree of Masters of Science at George Mason University

By

Ann K. Emery  
Bachelor of Arts  
University of Virginia, 2008

Advisor: Rodney Hopson, Professor  
College of Education and Human Development

Spring Semester 2014  
George Mason University  
Fairfax, VA

## Table of Contents

	Page
List of Tables .....	iii
List of Figures .....	iv
Abstract .....	2
Chapter 1. Introduction .....	1
Chapter 2. Literature Review .....	14
Chapter 3. Method .....	54
Chapter 4. Results .....	65
Chapter 5. Discussion .....	91
Appendix. Survey Instrument .....	124
References .....	135

## List of Tables

Table	Page
Table 1. Corresponding positive and negative experiences.....	71
Table 2. The evaluator's experience, knowledge, and soft skills.....	72
Table 3. Qualities of the evaluation process as a whole.....	80
Table 4. Methods, results, reports, and recommendations.....	83
Table 5. Project management: Timing and communication.....	86
Table 6. Collaboration (or lack of collaboration) with staff.....	88
Table 7. Red flags.....	90

## List of Figures

Figure	Page
Figure 1. Concept map.....	13
Figure 2. Timeline of prior research studies .....	15
Figure 3. Nonprofit evaluation encompasses a number of topics .....	22
Figure 4. Percentage of nonprofits engaged in evaluation.....	23
Figure 5. Percentage of nonprofits using a randomized control trial.....	28
Figure 6. Percentage of nonprofits working with external evaluators .....	34
Figure 7. Number of survey responses .....	55
Figure 8. The survey's population, sampling frame, and sample .....	56
Figure 9. Demographic characteristics of organizations represented in survey .....	57
Figure 10. Budget sizes of organizations in sample versus population .....	58
Figure 11. Regional distributions of organizations in sample versus population .....	59
Figure 12. States represented in sample.....	60
Figure 13. Skip pattern for two-part survey question .....	61
Figure 14. Percentage of nonprofits working with internal and/or external evaluators....	66
Figure 15. Nonprofit ratings of experiences with external evaluators .....	67
Figure 16. Reflections about emerging findings.....	68
Figure 17. Word counts for open-ended responses.....	69
Figure 18. Multiple explanations as to why evaluation experiences were negative.....	70
Figure 19. Multiple explanations as to why evaluation experiences were positive.....	71

## **Abstract**

### NONPROFITS' EXPERIENCES WITH EVALUATION

Ann K. Emery, M.S.

George Mason University, 2014

Thesis Chair: Dr. Rodney Hopson

The purpose of this study is two-fold: 1) to understand how many nonprofits are working with internal and/or external evaluators and 2) to learn how they describe their experiences working with evaluators. The study involved a secondary analysis of survey data from a national sample of 546 501(c)3 nonprofit organizations. I found that a minority of nonprofits reported working with a professional evaluator: sixteen percent worked with an internal evaluator and twenty-five percent worked with an external evaluator. Of these, more than two-thirds (sixty-nine percent) said the experience was *excellent* or *good*. When asked to describe why experiences with external evaluators were either positive or negative, nonprofits named six different types of reasons: 1) the particular evaluator's prior experience, knowledge, or soft skills; 2) impressions of the evaluation process as a whole; 3) specific aspects of the evaluation's methods, findings, reports, or recommendations; 4) project management issues, such as the project's

timeframe and communication; 5) collaboration—or a lack of collaboration—with nonprofit staff during the evaluation; and 6) “red flags,” for example, when the evaluation is used for fundraising or marketing rather than for learning and improvement. Taken together, these findings can help evaluators understand what types of approaches and project management styles are most likely to lead to satisfying and productive evaluation engagements with nonprofits.

## **Chapter 1. Introduction**

Nonprofit organizations are dedicated to solving social, economic, health, educational, and environmental problems that government institutions cannot or will not address. How can nonprofits ensure that their services are effective and efficient so that they are in a position to alleviate these problems? How can their funders—the government, foundations, corporations, and the public—have confidence that their donations are being used well? Evaluation is one strategy that allows nonprofits to manage and improve their own work and to demonstrate transparency to funders—improving their organization while proving that funds are being allocated as intended.

This paper explores nonprofit evaluation practices and capacity, with a focus on nonprofits' perspectives on the evaluation process. Chapter 1 provides background information about nonprofits, a brief overview of prior research, and the purpose of the present study (my goals, research questions, conceptual framework, methods, and validity). Chapter 2 provides an in-depth literature review that summarizes the past fifteen years of research. Chapter 3 outlines the methods used in the present study and Chapter 4 illustrates the findings from the present study. Finally, Chapter 5 discusses the implications of those findings, the limitations of this study, and suggestions for future research.



## **Introduction to Nonprofits**

Nonprofit organizations operate for charitable, educational, religious, or scientific purposes, or to prevent cruelty to children or animals. According to Carman, Fredericks, and Introcaso (2008), nonprofits are particularly valuable because they are dedicated to public charitable purposes. Nonprofits contribute to “relief of the poor, the distressed, or the underprivileged; advancement of religion; advancement of education or science; erecting or maintaining public buildings, monuments, or works; lessening the burdens of government; lessening neighborhood tensions; eliminating prejudice and discrimination; defending human and civil rights secured by law' and combating community deterioration and juvenile delinquency” (Internal Revenue Service, 2007, p. 4).”

There are 2.3 million nonprofit organizations in the United States (Roeger, Blackwood, & Pettijohn, 2011). States with the largest populations (such as California, New York, and Texas) are home to the greatest number of nonprofits while western states (such as Wyoming, Nevada, and Idaho) contain the fewest nonprofits (Carman et al., 2008). Although the U.S. Tax Code describes twenty-eight different categories of nonprofit organizations, a common feature of nonprofits is that they are exempt from Federal income taxes according to Section 501(c) of the U.S. Internal Revenue Code.

In 2010, nonprofits contributed \$804.8 billion to the United States economy, making up 5.5 percent of the country’s gross domestic product (Blackwood, Roeger, & Pettijohn, 2012). Public charities reported \$1.51 trillion in revenue, \$1.45 trillion in expenses, and \$2.71 trillion in assets in 2010, and private charitable contributions (such

as donations to religious congregations) totaled \$298.4 billion in 2010 (Roeger et al., 2011).

*Nonprofits Under a Microscope and the Emergence of the Outcomes Movement*

“Md. woman accused of stealing millions from D.C. nonprofit group,” read the November 24, 2013 headline from the Washington Post (Flaherty & Stephens, 2013). The article went on to describe how Ephonia Green, an administrative assistant making \$56,000 a year, appears to have embezzled more than \$5.1 million from the Association of American Medical Colleges (AAMC) between 2005 and 2013. Green stole nearly \$1.4 million by writing checks out to her own bridal shop. She allegedly stole another \$3.7 million by writing checks out to “Brookings Institute” rather than “Brookings Institution”—and then the money went straight into bank accounts she had created herself. She also “submitted small, bogus bills to cover the registration fees of staff members supposedly headed to conferences.” And how did Ephonia Green embezzle \$5.1 million without the AAMC’s knowledge? President Darrell Kirch merely offered this explanation: “Nonprofits have lacked some of the rigor that is enforced in for-profit organizations” (Flaherty & Stephens, 2013).

Regrettably, stories of nonprofit embezzlement are not new. In 2009, for example, an employee stole \$2 million from Youth Service America, a national nonprofit also based in Washington, DC (Flaherty & Stephens, 2013). Perhaps the most famous example is from the early 1990s, when the United Way of America’s (UWA) chairman received an anonymous note that CEO William Aramony was stealing from the organization. Amidst investigations from the Washington Post, Internal Revenue Service,

and the FBI, Aramony retired. He was later convicted for defrauding the United Way of more than \$1.2 million and was sent to prison. This scandal “damaged UWA’s credibility and fundraising efforts at local UWs and helped fuel a decline in trust in the nonprofit system in general” (Hendricks, Plantz, & Pritchard, 2008, p. 14-15).

Each new scandal puts nonprofits under a microscope in the public eye. In order to “prevent future abuses and rebuild trust,” the United Way launched its *outcomes movement*, which emphasized transparency and accountability and cost the United Way more than \$2.4 million between 1995 and 2000 (Hendricks et al., 2008, p. 14-15). Part of this movement involved the creation of five resources about nonprofit evaluation. The most well-known resource is *Measuring Program Outcomes: A Practical Approach* (United Way of America, 1996), a 170-page manual with lead authorship by Harry Hatry, a performance measurement expert from the Urban Institute (Hendricks et al., 2008, p. 21-22). This outcomes movement persisted into the 2000s and “with this pressure from federal and United Way funding” (Behrens & Kelly, 2008, p. 39) nonprofits nationwide began to develop logic models, collect data, and report on outcomes. A full literature review about trends in outcomes and evaluation is provided in Chapter 2.

Today, the terms *outcomes measurement*, *performance measurement/management*, *program evaluation*, and *evaluation* refer to similar processes in which nonprofits collect and analyze data about the effectiveness of their work. *Outcomes measurement* means tracking “what happens to participants after they receive your services” such as “what participants know, think, or can do; or how they behave; or what their condition is, that is different following the program” (United Way of America,

2004, p. 2). *Performance measurement* is an ongoing cycle in which nonprofits routinely collect and analyze data. This process is often by nonprofit staff rather than by evaluation consultants and is intended to be a regular part of organization management.

*Performance management* takes this ongoing cycle one step further: the expectation is that nonprofits are using that information to make decisions about future programming and strategies. *Program evaluation*, perhaps the most common term today, is “a social science activity directed at collecting, analyzing, interpreting, and communicating information about the workings and effectiveness of social programs” (Rossi, Lipsey, & Freeman, 2004, p. 2). Recognizing that much nonprofit work falls outside the boundaries of a program—advocacy initiatives, organization-level data, a foundation’s portfolio-level data, and others—many evaluators have transitioned into calling this process *evaluation* rather than *program evaluation*.

### **Problem Statement**

Nonprofits have been engaging in evaluation since at least the early 1990s. To date, however, there have not been any systematic efforts to collect nonprofits’ feedback on the evaluation process or to share that feedback with evaluators. While some evaluators may have decades of experience working with nonprofits, others are embarking on their first project with a nonprofit and may not know what to expect—or what common pitfalls should be avoided. Similarly, even experienced evaluators may not realize which approaches have been especially well-received by their nonprofit partners or which approaches should be avoided in future partnerships.

This project is an opportunity to carefully review nonprofits' feedback on factors that help or hinder the evaluation process. In future professional endeavors I will seek opportunities to share these findings with evaluators so that they can improve their partnerships with nonprofits.

### **The Current Study**

I have outlined the current study's goals, conceptual framework, research questions, methods, and validity according to Maxwell's (2005) interactive model for qualitative research.

#### *Goals*

According to Maxwell's (2005) interactive design, *goals* specify why a study is worth conducting, what issues the study can clarify, and what practices and policies the research wants to influence. Goals can be personal, professional, or for the pursuit of new knowledge (K. Sheridan, personal communication, September 12, 2011). I have three goals in pursuing this research: for funders, for nonprofits, and for evaluators. First, and most importantly, my goal is to learn about nonprofits' perspectives on their experiences working with evaluators. To my knowledge, no prior studies have focused on nonprofits' views, nor have any sought to gather feedback on evaluators or "evaluate" the evaluator. This first goal is an attempt to generate new knowledge for the field of nonprofit evaluation. Second, after the completion of this study, I want to share this nonprofit feedback with the evaluation community. This second goal is both a personal and professional goal in which I will aim to influence and improve evaluation practices. Third, I seek to improve funders' knowledge of what types of evaluation nonprofits are

undertaking—and how much capacity for evaluation really exists, or needs strengthening. I anticipate sharing key findings from the literature review with foundation audiences during upcoming panels, conference presentations, white papers, and other outlets. This goal is a personal and professional interest, and is a vehicle for sharing new knowledge with a broader group of stakeholders so that practice and policy may be improved.

### *Conceptual Framework*

A *conceptual framework* describes what researchers think is going on with the issues, settings, or people that will be studied; what theories, beliefs, and prior research findings will guide the study; and what personal experiences the researcher will draw upon for understanding the research topics (Maxwell, 2005). My conceptual framework about nonprofit evaluation is guided by the following influences: my own career background as an internal evaluator in a community-based nonprofit; my career background as an external evaluator to both nonprofits and foundations; conversations with friends and colleagues working in the nonprofit sector about evaluation (e.g., over coffee, during conference sessions, etc.); my experience researching and writing *State of Evaluation 2012*, the dataset from which this secondary analysis is pulled; and my literature review of the past fourteen years of research about nonprofit evaluation (described fully in Chapter 2).

### *Research Questions*

*Research questions* specify what, specifically, will be understood through the study (Maxwell, 2005). This current study is guided by a two-part research question:

1. How many nonprofits are working with internal and/or external evaluators?

2. How do nonprofits describe their experiences with evaluation? (For example, what types of phrases or language are most commonly used? How many distinct categories emerge? Can categories be grouped into a hierarchy? How many nonprofits fall into each of the categories?)

### *Methods*

According to Maxwell's (2005) interactive design, *methods* describe what the researcher will do while conducting the study. This study is a secondary analysis of closed-ended and open-ended survey responses from an existing national survey of nonprofits.

In the larger study, 546 nonprofits completed Morariu et al.'s (2012) State of Evaluation survey. The full survey included thirty-eight questions about evaluation practices and capacity, of which twenty-five questions were closed-ended and five questions were open-ended. The full survey instrument is provided in the appendix. When conducting the State of Evaluation 2012 research, Morariu et al. (2012) contacted GuideStar ([www.GuideStar.org](http://www.GuideStar.org)) to purchase contact information for U.S.-based 501(c)3 Public Charities. The researchers were interested in organizations that had updated their IRS Form 990 in 2010 or more recently and had provided a contact name and email address. A total of 38,789 organizations met the criteria. In August 2012, Morariu et al. (2012) sent an initial email invitation to a random sample of 20,000 organizations, followed by three additional reminder emails to representatives who had not yet responded. The survey was open for twenty-five days. Throughout the survey administration period, logistical and technical issues prevented the researchers from

reaching all 20,000 organizations. For example, several thousand surveys could not be delivered because the receiving email server was busy and not receiving messages; because the receiving email inbox was full; because the receiving organization's firewall blocked the message as spam; or because the email address was incorrect due to a typo or other error. In several cases, Morariu et al. (2012) received automated messages that indicated that the organizations were no longer open. Out of 20,000 organizations, 13,070 organizations received the survey and 546 responses to the survey, for a response rate of four percent.

The current study is a deeper analysis of two survey items, Question 17 and Question 18/19. Question 17 asked, "If your organization worked with an external evaluator in 2011, what was your experience?" Respondents could select *horrible*, *poor*, *mixed*, *good*, *excellent*, *we did not work with an external evaluator*, or *I don't know*. Respondents who indicated that their experience was *horrible*, *poor*, or *mixed* were led to Question 18, an open-ended item that asked, "You indicated that you worked with an external evaluator in 2011 and it wasn't a great experience. Why do you think the experience wasn't positive?" Respondents who indicated that their experience was *good* or *excellent* were led to Question 18, an open-ended item that asked, "You indicated that you worked with an external evaluator in 2011 and it wasn't a great experience. Why do you think the experience was positive?" Nonprofits that responded *we did not work with an external evaluator* or *I don't know* were automatically skipped ahead to the next survey question. I coded the responses and found fifteen different codes for Question 18 (for nonprofits that indicated having a negative evaluation experience) and twenty-two



different codes for Question 19 (for nonprofits that indicated having a positive evaluation experience). Then, I grouped the codes into six larger categories based on whether the survey respondent was describing the external evaluation consultant versus the broader evaluation process and whether the respondent was describing evaluation skillsets or techniques that can be learned in a classroom or training setting versus on-the-job experience.

Multiple research projects have sought to better understand how and why nonprofits measure the results of their work (e.g., Alaimo, 2008; Carman, 2004; Carman, 2007; Carman, 2011; Carman & Fredericks, 2008; Carman & Fredericks, 2009; Carman et al., 2008; Carman & Millesen, 2004; Fine et al., 2000; Hoefler, 2000; Morariu et al., 2012; and Reed & Morariu, 2010, among others). This study's design has both similarities and differences to prior research on nonprofit evaluation practices.

The larger study from which the open-ended survey items are pulled, State of Evaluation 2012, is similar to other studies. Most prior research on nonprofit evaluation has involved survey research projects (e.g., Carman, 2004; Carman, 2007; Carman & Fredericks, 2008; Carman & Millesen, 2004). Less often, research on nonprofit evaluation has involved interviews, case studies, and/or document review (e.g., Alaimo, 2008; Carman, 2011). The larger study is therefore similar to many prior studies because it employs a survey research design. However, most prior research survey samples have involved nonprofits from a specific geographic area, such as nonprofits from New York (Carman, 2004; Carman, 2007; Carman, 2011) or Indiana (Carman & Fredericks, 2008). To date, there are three national studies of nonprofits in the United States: a study by Fine

et al. (2000); Johns Hopkins University's Listening Post Project (Salamon, Geller, & Mengel, 2010); and the State of Evaluation research project (Reed & Morariu, 2010; Morariu et al., 2012). The larger study is somewhat unique, therefore, in its national-level focus.

The current study, an analysis of open-ended survey items, is also somewhat unique when compared to prior research on nonprofit evaluation. The focus of the current study is on nonprofits' descriptions of working with external evaluation consultants. To date, no studies have focused on this topic. The analysis techniques are also unique. In the current study, I coded open-ended survey responses, while prior qualitative studies have summarized key findings from interviews with nonprofit staff (e.g., Alaimo, 2008; Carman, 2004; and other interview studies).

The methods and data analysis process are described fully in Chapter 3 and the results of the coding procedures are described fully in Chapter 4. In Chapter 5, I connect the codes found in this study to prior research.

### *Validity*

Maxwell (2005) defines *validity* as how the study's results and conclusions can be wrong; for example, what types of alternative interpretations are also plausible. A primary validity threat associated with this study is that I could misinterpret responses during the coding process. That is, my biases from my prior experiences with nonprofits could lead me to miss an important pattern in the information or to overemphasize or underemphasize certain phrases in the responses. To counter this validity threat, I shared and discussed preliminary findings with the committee chair, Dr. Hopson, to help

determine whether I was interpreting the responses accurately and consistently. I also discussed the codes with three additional committee members during the oral defense.

Figure 1 is a concept map that summarizes the current study's goals, conceptual framework, research questions, methods, and validity.

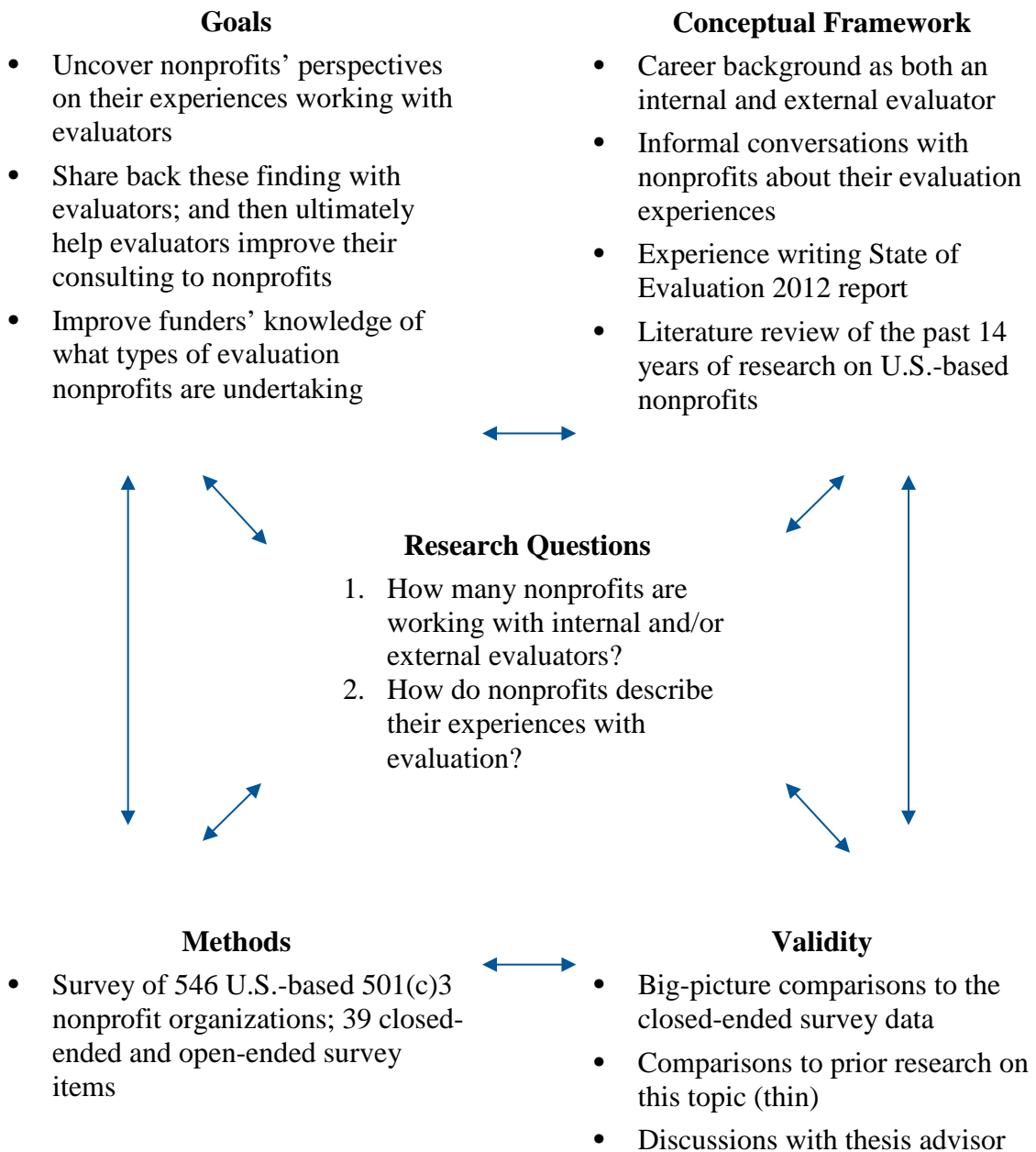


Figure 1. Concept map

## **Chapter 2. Literature Review**

Multiple research projects have sought to better understand how and why nonprofits measure the results of their work. This literature review first describes those studies and then highlights key findings.

### **Prior Research Studies on Nonprofit Evaluation**

Over the past fourteen years, numerous studies on nonprofit evaluation have been conducted. Existing research primarily consists of self-report surveys with nonprofit representatives. Additionally, some researchers have conducted interviews, reviewed evaluation documents, and compiled case studies. While some of these studies are peer reviewed, most are conference papers and white papers.

The following timeline displays when these studies were published. As illustrated in the timeline, the first studies appeared in 2000, likely as a result of the new push for outcomes management by the United Way and several other foundations. Several additional studies were released in 2003 and 2004. Since 2007, between one and three new studies have been released each year, demonstrating a steady interest in nonprofit evaluation by researchers as well as the funders who provide resources to conduct those studies.

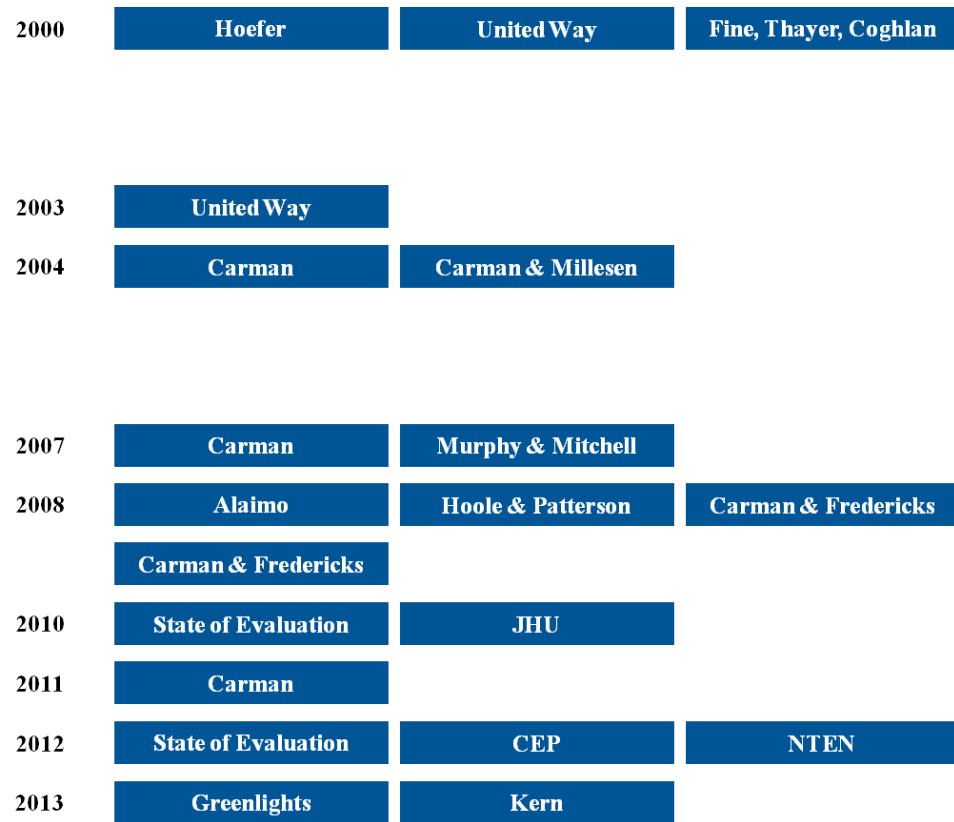


Figure 2. Timeline of prior research studies

### *Surveys, Interviews, and/or Case Studies at the Local Level*

A number of studies have focused on nonprofits in a specific geographic area (e.g., nonprofits in Illinois or nonprofits in Ohio). For example:

- In 2000, Hoefler collected ninety-one mail surveys from direct services organizations in Dallas. The analysis examined whether evaluations were conducted, the types and designs of evaluations conducted, the types of data collected, and how findings are used. The study also focused on funding sources

for evaluation and whether nonprofits work with internal and/or external evaluators.

- United Way of America also conducted survey research in 2000 to determine the extent to which United Way-funded nonprofits were engaging in outcomes measurement. United Way collected 298 mail surveys from direct services organizations and the sample was drawn from six different geographic regions of the country.
- In 2003, United Way of America surveyed thirty-three national-level organizations (e.g., Big Brothers Big Sisters and the Boy Scouts of America) and four accreditation organizations. The goal of this study was to learn about how national organizations were supporting their affiliates and chapters in outcome measurement. United Way also provided narrative text that described each organization's evaluation practices.
- In the mid- to late-2000s, Carman published a number of studies related to her dissertation research with nonprofits in New York. Each study utilized a slightly different sample and had a different purpose, as will be described next.
- In 2004, for example, Carman and Millesen collected 305 surveys (178 from nonprofits in New York and 127 from nonprofits in Ohio). Their analysis focused on nonprofits' evaluation activities, such as the types of data collected and the percentage of nonprofits using logic models. They also measured nonprofits' resource needs and barriers to evaluation, such as their staffing limitations.

- In 2007, Carman published more results from her research with nonprofits in New York. This paper sought to determine whether there were differences between three types of nonprofits: social services organizations, organizations serving people with developmental disabilities, and community development organizations. She compared the types of data collected, the funding sources for evaluation, and the evaluation staffing duties at each type of organization.
- Alaimo (2008) conducted forty-two interviews with human services organizations in Atlanta and supplemented the interview data with case studies about two organizations' evaluation practices.. He focused on the executive director's role in evaluation capacity building as well as the tension between the external push for accountability and the internal pull for learning.
- Carman and Fredericks (2008) collected 189 surveys from nonprofits in Indiana. This study covered a breadth of topics, including the types of evaluation activities that nonprofits are engaging in, the types of data they collect, and how they use data. The authors also completed a factor analysis and found that nonprofits think about evaluation in three different ways: as a tool to promote their work to external audiences, as a tool for managing the organization, and as a resource drain and distraction.
- Hoole and Patterson (2008) shared three case studies that focused on organizational learning practices that were in place in each of the organizations—in other words, how the organizations use evaluation as a process of continuous improvement and growth in which evaluation is embedded within the



organization's systems, structure, leadership, and culture. This was the first time that learning practices were showcased; prior studies focused more narrowly on data collection practices (e.g., how many nonprofits were conducting focus groups or administering pre- and post-tests) and on evaluation designs (e.g., how many nonprofits were collecting data from comparison or control groups).

- In 2009, Carman and Fredericks published more findings from the 189 surveys they collected from nonprofits in New York. They also conducted twenty-six follow-up interviews. This study focused on nonprofits' barriers to evaluation: whether the survey respondents and interviewees felt they had adequate time, expertise, and leadership support to engage in evaluation practices.
- In 2011, Carman published findings from thirty-one interviews with nonprofits in New York. She supplemented the interview data with a document review of evaluation reports, brochures, data collection forms, and other materials that described the nonprofits' evaluation practices. The analysis focused on inter-organizational factors such as nonprofits' relationships with funders and their competition for resources with other nonprofits.
- In 2012, the Nonprofit Technology Network collected 398 surveys and conducted six telephone focus groups to study nonprofits' use of data for decision-making. This study was not limited to evaluation data; instead, the survey asked nonprofits to describe their use of financial and operations data; marketing, communications, and fundraising data; and programmatic and outcomes data.

- In 2012, Brock, Buteau, and Herring (2012) surveyed 177 nonprofits from the Center for Effective Philanthropy’s Grantee Voices panel. The researchers sought to understand whether nonprofits felt it was important to evaluate their work—and found that four in five nonprofits (81%) believe nonprofits should demonstrate the effectiveness of their work through performance measurement and evaluation.

### *National Studies on Nonprofit Evaluation*

To date, there are three national studies of nonprofits in the United States: a study by Fine et al. (2000); Johns Hopkins University’s Listening Post Project (Salamon et al., 2010); and the State of Evaluation research project (Reed & Morariu, 2010; Morariu et al., 2012).

In 2000, Innovation Network’s Fine et al. (2000) collected 178 mail surveys from a nationwide sample of direct services organizations. Researchers also conducted 40 interviews and developed four in-depth profiles. The study was primarily descriptive, showing why and how nonprofits conduct evaluations. They also examined whether evaluations with high levels of stakeholder involvement were more or less likely to be perceived as useful and credible. The first nationwide survey of its kind, this study offered new insights on how nonprofits are (and are not) utilizing evaluation. However, the survey sample was not representative—organizations were nominated by staff at foundations and management support organizations and by members of the American Evaluation Association’s Independent Consultants Topical Interest Group.

In 2010, Johns Hopkins University's Listening Post Project surveyed a nationwide sample of 417 nonprofits. Amidst the buzz around the Obama administration's Social Innovation Fund, the research focused on nonprofits' use of innovative practices as well as how they have adopted performance measurement. Again, this sample was not representative of all nonprofits nationwide. Only four types of nonprofits were included in the sample: nonprofits that focus on children and family services, elderly housing and services, community and economic development, and the arts.

Also in 2010, Innovation Network's Reed and Morariu launched the State of Evaluation research project and Morariu et al. launched the second iteration of the project in 2012. One thousand seventy-two nonprofit representatives completed the 2010 survey and 546 nonprofits completed the 2012 survey. The research is cross-sectional; that is, a similar (but not identical) group of nonprofits are randomly selected to participate in each wave of the study. The State of Evaluation project has spurred several related studies:

- In 2013, Greenlights for Nonprofit Success, an evaluation and research consulting group, used the State of Evaluation questionnaire as the basis for their survey research with nonprofits in central Texas (Greenlights for Nonprofit Success, 2013). The study sought to determine whether nonprofits in central Texas used similar evaluation practices as nonprofits nationwide.
- In 2013, Master's student Trisa Kern examined a subset of State of Evaluation 2012 survey data. In particular, she examined 103 organizations located in the Midwest and supplemented the survey data with 16 telephone interviews. She

sought to determine which characteristics of nonprofits correlate with evaluation use and found that “organizations receiving funds from external sources, and those serving a larger region, tend to use evaluation data in more ways” (Kern, 2013, p. i).

- Roger Miranda is currently examining a subset of the State of Evaluation 2012 dataset. He is contributing to several pro bono evaluations with nonprofits in Canada. He will examine the State of Evaluation 2012 dataset to find survey respondents that are similar to the Canadian organizations with which he is working. Then, he aims to develop a rubric or checklist of evaluation practices and capacities. He plans to compare the Canadian nonprofits’ current evaluation infrastructure to this rubric and help those nonprofits strengthen their existing practices (R. Miranda, personal communication, September 3, 2013).

### **Key Patterns from Prior Research**

The term *nonprofit evaluation* involves a number of research designs, data collection practices, and dissemination techniques. In particular, when discussing nonprofit evaluation, nonprofits, evaluators, and funders may refer to any of the topics shown in the following figure, such as the percentage of nonprofits engaging in evaluation, whether they create logic models, the types of research designs used, how they staff their evaluation with internal and/or external evaluation professionals, and how they budget for evaluation, among other topics.



Figure 3. Nonprofit evaluation encompasses a number of topics

In the following section, I describe: (1) how many and what types of nonprofits are engaging in evaluation; (2) evaluation practices used, such as which research designs and data collection methods are most common; (3) staffing for evaluation, such as whether nonprofits work with internal or external evaluators; (4) evaluation use, such as how findings are disseminated and used for decision making; (5) funding for evaluation; (6) barriers to conducting and using evaluation; (7) benefits of evaluation; and (8) nonprofits' perspectives on the meaning and purpose of their evaluation efforts.

## How Many Nonprofits are Engaging in Evaluation

Most nonprofits are engaging in evaluation of some kind. This bar chart illustrates the percentage of nonprofits that reported being involved in evaluation.

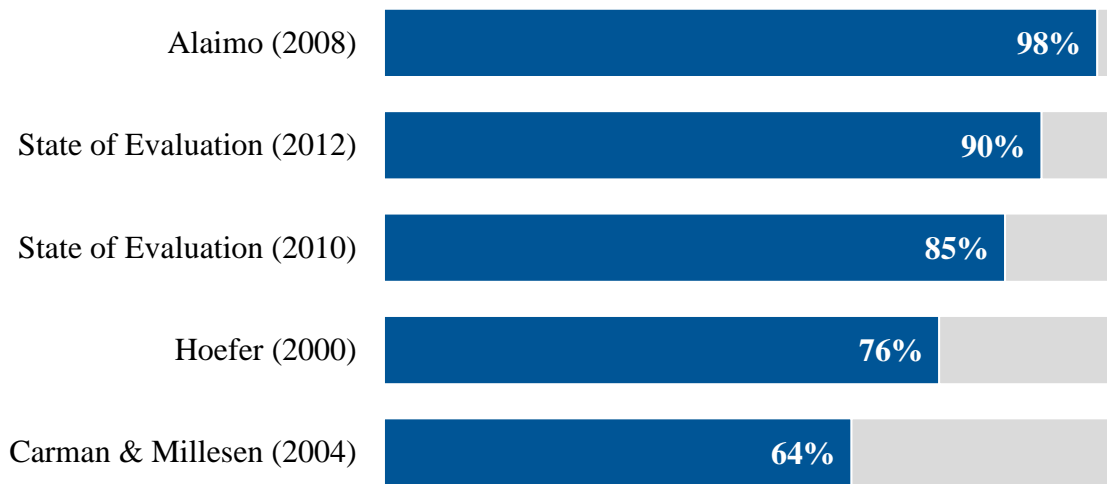


Figure 4. Percentage of nonprofits engaged in evaluation

Variations in percentages are likely due to the study's time period (i.e., the year in which it was conducted), the study sample, how evaluation was defined, and how the question was asked in the survey or interview. For example, in 2000, seventy-six percent of organizations indicated that the largest program in their organization had been evaluated within the past two years (Hoefler, 2000). In 2008, forty-one of the forty-two organizations interviewed by Alaimo (2008) were engaged in evaluation. In the State of Evaluation 2012 survey, ninety percent of nonprofits nationwide reported that they

evaluated some aspect of their work within the past year (Morariu et al., 2012, p. 2). This represents a slight increase from the 2010 State of Evaluation survey in which eighty-five percent of nonprofits reported engaging in evaluation (Reed & Morariu, 2010, p. 2).

However, it is unlikely that such a high percentage of nonprofits are engaged in full, in-depth evaluations. Instead, it is more likely that nonprofits are defining evaluation loosely as collecting and analyzing some type of data about program implementation or outcomes. For example, in 2004, sixty-four percent of the organizations reported conducting “formal program evaluations” but only forty-one percent reported using a performance measurement system (Carman & Millesen, 2004, p. 15). Carman and Fredericks (2008) found similar results: while ninety percent of nonprofits engaged in some type of evaluation, fifty-five percent regularly conduct formal evaluations and forty-six percent used a performance measurement system. In that same study, eighteen percent of nonprofits said “we go out of our way to evaluate all of our programs” while the remaining respondents said they “make a concerted effort to evaluate [their] programs” (forty-six percent) or that they “do some evaluation” (twenty-six percent). Five percent of respondents engage in “very little evaluation” and five percent did not evaluate any of their programs (Carman & Fredericks, 2008, p. 56). In 2012, the Nonprofit Technology Network found that ninety-nine percent of nonprofits “track some sort of metrics,” such as financial and operations data, but only fifty percent of nonprofits reported tracking data about client outcomes (Nonprofit Technology Network, 2012, p. 2). In a survey of 156 nonprofits in central Texas, ninety-three percent of respondents reported that they engaged in evaluation, although only fifty-three percent track medium-

term outcomes and only thirty-seven track long-term outcomes (Greenlights for Nonprofit Success, 2013, p. 5). Taken together, these findings suggest that the depth and type of evaluation activities varies widely among nonprofit organizations.

Certain types of organizations are more likely to engage in evaluation than others. Two factors have consistently been found to influence the likelihood of a nonprofit engaging in evaluation: budget size and organization age. For instance, multiple studies have found that nonprofits with larger budgets are more likely to engage in evaluation than nonprofits with smaller budgets. Larger nonprofits (defined by the researchers as organizations with annual budgets of \$5 million or more) were more likely than smaller organizations (those with annual budgets of \$500,000 or less) to evaluate their work: ninety-four percent of large nonprofits engaged in evaluation compared to seventy-eight percent of smaller nonprofits (Morariu et al., 2012, p. 2). In the survey of nonprofits in central Texas, researchers also found differences between small and large nonprofits. For example, seventy-two percent of large organizations collected data on long-term outcomes (compared to twenty-nine percent of small organizations). In addition, sixty percent of large organizations collected data on medium-term outcomes, compared to forty-six percent of small organizations (Greenlights for Nonprofit Success, 2013, p. 6). Older organizations (defined as organizations that are more than twenty years old) were also more likely to engage in evaluation than younger organizations (organizations that were one to five years old): ninety percent of older organizations engaged in evaluation compared to eighty percent of younger organizations (Morariu et al., 2012, p. 2).

### **Logic Models, Evaluation Designs, and Data Collection Methods**



### *Logic Models*

Logic models are diagrams or flow charts that describe how a nonprofit's activities lead to desired outcomes over time. Logic models can serve several purposes. For example, the process of building a logic model often builds consensus among nonprofit staff about the program's longer-term role and purpose. They also serve as external communication tools; logic models are required in virtually all grant applications in order to display the program's goals and objectives. During evaluation planning, logic models serve the basis of an evaluation plan because they help nonprofits and evaluators determine what should be evaluated.

Logic models gained popularity during the United Way's outcomes movement of the 1990s and 2000s. By 2004, twenty percent of nonprofits designed logic models (Carman & Millesen, 2004, p. 15). Similar results were found in 2007 and 2008; Carman (2007) reported that seventeen percent of nonprofits designed logic models and Carman and Fredericks (2008) reported that twenty-three percent of nonprofits designed logic models. By 2012, sixty percent of organizations had a logic model, but fewer organizations (forty-one percent) had created or revised that logic model or theory of change within the past year (Morariu et al., 2012, p. 6). Morariu et al. (2012) also found differences by organization size: forty-five percent of large organizations had revised their logic model or theory of change within the past year (down from fifty-six percent in 2010) and thirty percent of small organizations had revised their logic model within the past year (down from thirty-four percent in 2010). These findings suggest that there is room to grow; how can logic models help nonprofits articulate their goals and objectives

if fewer than two-thirds of nonprofits are using logic models and fewer than half of nonprofits are updating them regularly?

### *Evaluation Designs and Approaches*

There are a number of evaluation designs and approaches, from collecting before and after measures to gauge participant progress to engaging in a developmental evaluation. Outcomes evaluations are consistently used by the most nonprofits, perhaps due to the United Way's outcomes movement and accompanying handbooks, guidebooks, and training that advocated for this approach. As early as 2000, eighty-two percent of nonprofits reported engaging in an outcomes evaluation (Hoefler, 2000); by 2012, seventy-nine percent of nonprofits reported engaging in an outcomes evaluation (Morariu et al., 2012). Pretest-posttest designs are also popular: forty-nine percent of nonprofits used a pretest-posttest design in 2000 and sixty-five percent of nonprofits used this design in 2012 (Hoefler, 2000; Morariu et al., 2012). Most recently, nonprofits also reported using formative evaluations (seventeen percent of survey respondents) and summative/retrospective evaluations (thirty-eight percent) as well as developmental evaluation (nineteen percent) and its close relative, strategic learning (twenty percent; Morariu et al., 2012). The high percentage of nonprofits reporting engaging in developmental and strategic learning approaches is likely an indicator of low evaluation capacity. That is, it is likely that nonprofits do not fully understand these terms and are not actually engaged in this type of evaluation.

While nonprofits are most likely to engage in outcomes evaluation, they are least likely to engage in an evaluation that involves control or comparison data, such as a

randomized control trial. This bar chart illustrates how few nonprofits reported using a randomized control trial during their evaluation.

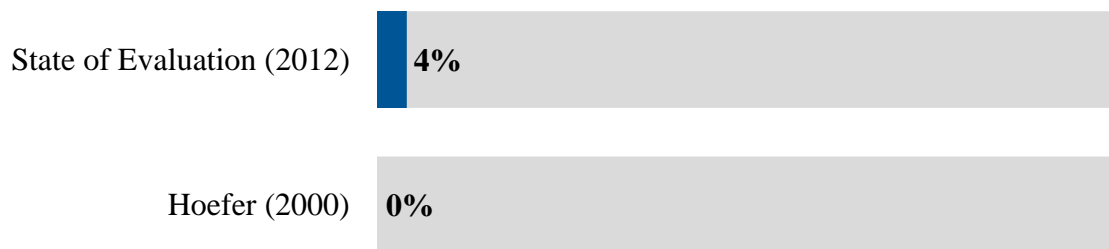


Figure 5. Percentage of nonprofits using a randomized control trial

For example, based on his conversations with nonprofit directors, Hoefer (2000) estimated that none of the nonprofits in his study were involved in randomized control trials. Even by 2012, only four percent of nonprofits had been involved in a randomized control trial (Morariu et al., 2012). Nonprofits are slightly more likely to use evaluation designs that involve comparison groups, such as quasi-experimental designs. The percentage of nonprofits using control or comparison data has varied between six percent (in the 2012 national study by Morariu et al.) and seventeen percent (in the 2000 study of nonprofits in Dallas by Hoefer). Carman (2007) found differences by organization type. That is, fifteen percent of social services organizations collected control or comparison data, compared to ten percent of organizations serving people with developmental disabilities and seven percent of community development organizations.

#### *Types of Data Collected*

Most commonly, nonprofits collect information about activities and outputs (such as the number of people served and their demographic characteristics). Carman (2004) found that sixty-nine percent of nonprofits collected data on program activities and Carman and Fredericks (2008) found that sixty percent of nonprofits *regularly* collected data on program activities or outputs. Furthermore, eighty-nine percent of nonprofits collected data on the number of people they served and seventy-seven percent collected demographic information about those participants (Carman & Fredericks, 2008).

Nonprofits also tend to collect data on participant satisfaction: depending on the research study, between sixty-seven percent and seventy-two percent of nonprofits collect consumer satisfaction data (Carman & Millesen, 2004; Carman & Fredericks, 2008). Morariu et al. (2012) found that medium and large organizations were more likely to use feedback forms than small organizations (eighty-seven percent of medium and large organizations compared to seventy-six percent of small organizations).

Slightly fewer organizations collect outcomes data. In 2004, sixty-three percent of nonprofits gathered data on program outcomes or results, compared to a similar percentage—sixty-two percent—in 2008 (Carman & Fredericks, 2008; Carman & Millesen, 2004). However, only fourteen percent of organizations engaged in long-term follow-up with clients after they have stopped using the organization's services (Morariu et al., 2012). Nonprofits have described measuring long-term outcomes, in particular, “as the ‘holy grail’ of data-based decision-making” and the Nonprofit Technology Network concluded that “it’s unclear that this level of measurement is practical or even possible for many nonprofits” (Nonprofit Technology Network, 2012, p. 8).

These findings demonstrate that nonprofits are most likely to pursue relatively easy forms of data collection—counting the number of people served and their demographic characteristics—compared to more sophisticated forms of data collection—tracking those program participants after they have completed the program.

#### *Data Collection Sources*

Evaluation data can come from a variety of sources, including records, internal tracking forms, and surveys, among many others. In 2000, Fine et al. found that the most common data collection method was a review of program documents (eighty-four percent of survey respondents). Nonstandardized (fifty-four percent) and standardized (thirty-four percent) instruments are also used (Hoefler, 2000). Roughly half of nonprofits collect surveys or other questionnaires: Carman and Fredericks (2008) reported that approximately half of nonprofits used mail surveys and Fine et al. (2000) reported that forty-two percent of nonprofits collected written questionnaires. In addition, twenty-eight percent of nonprofits conducted telephone surveys (Carman & Fredericks, 2008) and nineteen percent administered test of knowledge or skills (Fine et al., 2000).

Qualitative data sources, such as data from interviews and focus groups, are also common. Interviews, in particular, are used by one-half to two-thirds of nonprofits. For example, Fine et al. (2000) found that two-thirds of nonprofits (sixty-six percent) conduct individual interviews and Carman and Fredericks (2008) found that fifty-nine percent of nonprofits conduct face-to-face interviews when they collect data. Interviews are used more often in medium and large organizations compared to small organizations (forty-

three percent of small organizations, fifty percent of medium organizations, and forty-nine percent of large organizations; Morariu et al., 2012).

Focus groups are used by one-quarter to one-third of nonprofits. Fine et al. (2000) found that one-third of nonprofits (thirty-three percent) conduct focus groups or group discussions and Carman and Fredericks (2008) found that twenty-nine percent of nonprofits conduct focus groups. Focus groups are more common in larger organizations; Morariu et al. (2012) reported that eighteen percent of small organizations, twenty-eight percent of medium organizations, and thirty-eight percent of large organizations use focus groups.

A variety of other data sources are used. Between fifty and seventy percent of nonprofits observe and record program activities (Carman & Fredericks, 2008; Fine et al., 2000). More than half of nonprofits collect narrative or anecdotal data (Carman & Millesen, 2004). Collecting anecdotal data refers to the practice of keeping mental notes of participant success stories (i.e., “cherry-picking” the sample), which contrasts with research methods such as interviews of focus groups that would systematically collect and analyze data from participants who have had both successful and unsuccessful experiences in the nonprofit’s program. This approach was jokingly named the “anecdote as evidence” model by the Nonprofit Technology Network (Nonprofit Technology Network, 2012, p. 9).

Some nonprofits, especially larger organizations, compile case studies. Thirteen percent of small organizations reported using case studies, compared to twenty-four percent of medium organizations and thirty-two percent of large organizations (Morariu

et al., 2012). Overall, two-thirds (sixty-seven percent) of survey respondents indicated that they used a mix of quantitative and qualitative data collection methods in their recently conducted evaluations and eighty-nine percent of interview respondents reported using both quantitative and qualitative data collection methods (Fine et al., 2000).

Researchers from several of these studies have suggested that these numbers are likely inflated. Carman (2007) commented, "Given that so many organizations reported that they were gathering data using written tools, interviews, and observation, these findings suggests that it is very likely that most of the data collection that community-based organizations are doing is being done in the context of providing services, as opposed to data collection for the expressed purpose of program evaluation and performance measurement" (Carman, 2007, p. 69). Morariu commented that is it unlikely that nineteen percent of nonprofits are actually engaged in developmental evaluation since so few evaluators actually practice developmental evaluation approaches (Morariu, 2012, personal communication). Patton described how individuals often confuse developmental evaluation with development evaluation (evaluation in developing countries) or developmental psychology (measuring changes in individuals over time; Patton, 2013). Furthermore, Carman and Fredericks (2008) described how nonprofits are more likely engage in reporting, regulatory, and monitoring activities than in management strategies or evaluation and performance measurement. For example, ninety-four percent of nonprofits produced reports for their board of directors (described by the authors as a reporting activity) and eighty-six percent conduct financial audits (a regulatory activity), but only fifty-five percent regularly conduct formal evaluations of

their programs (Carman & Fredericks, 2008). Clearly, nonprofits' understandings about evaluation data—as opposed to auditing, reporting, regulatory, or monitoring data—varies widely.

### **Staffing for Evaluation**

A variety of individuals can be tasked with evaluation duties: professional evaluators who work within the nonprofit (internal evaluators); professional evaluators who work for universities, consulting firms, or as independent consultants (external evaluators); or nonprofit staff such as the Executive Director who take on evaluation duties in addition to their other responsibilities.

#### *Internal Evaluators*

As the demand for evaluation increases, more nonprofits are hiring full-time professional evaluators. These embedded internal evaluators may have a variety of titles: evaluation specialist, quality assurance specialist, or director of research. They may be tasked with a variety of responsibilities, from training staff about evaluation to writing evaluation reports for funders to facilitating meetings about the evaluation's findings. Carman (2007) and Carman and Fredericks (2008) found that between four and eight percent of nonprofits had a dedicated staff member who was responsible for collecting evaluation data. Large organizations, in particular, are continuing to hire full-time internal evaluators. Morariu et al. (2012) found that fifty-three percent of large organizations had a full-time employee dedicated to evaluation (up from thirty-one percent in the State of Evaluation 2010 survey). In comparison, only nine percent of



small organizations had a full-time employee dedicated to evaluation, the same percentage as in the State of Evaluation 2010 survey.

*External Evaluators*

The percentage of nonprofits that work with external evaluation consultants is less clear, as the percentages vary widely based on how the question is framed in the research study. The following bar chart displays this range in findings from prior studies.

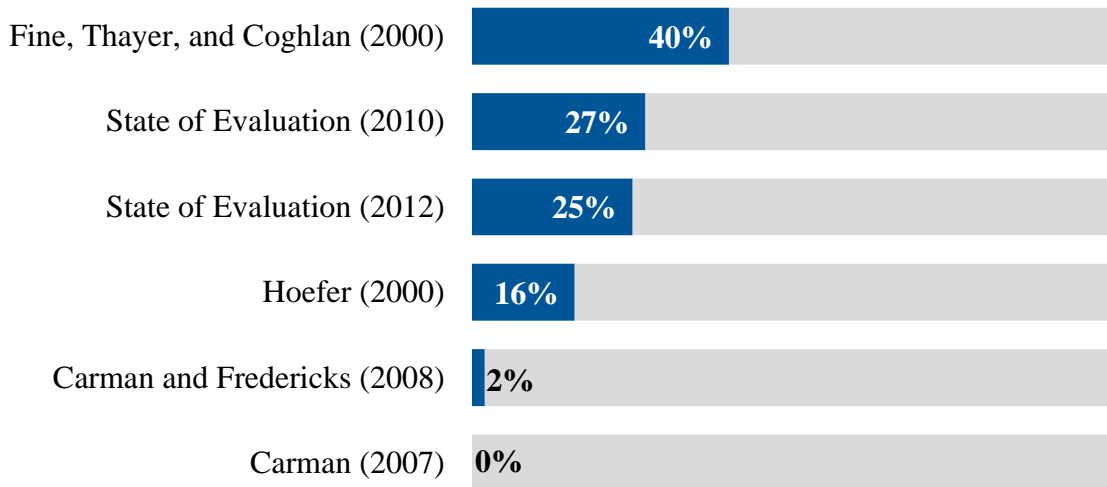


Figure 6. Percentage of nonprofits working with external evaluators

For example, zero percent of social services organizations reported that the primary responsibility for conducting an evaluation lies with an external evaluator in the study conducted by Carman (2007) and two percent of nonprofits reported that the primary responsibility for conducting an evaluation lies with an external evaluator in the

study conducted by Carman and Fredericks (2008). Forty percent of nonprofits said external contractors were responsible for leading their evaluation efforts (Fine et al., 2000) and sixteen percent only used outside evaluation expertise (Hoefer, 2000). In 2010, Reed and Morariu found that twenty-seven percent of nonprofits worked with an external evaluator the prior year. Large organizations are more likely to work with external evaluators than small organizations (forty-two percent versus seventeen percent; Morariu et al., 2012), presumably because large organizations are also more likely to evaluate their work in the first place.

#### *Combination of Internal and External Evaluators*

In some cases, nonprofits are working with both internal and external evaluators. In 2000, Hoefer found that nearly three-fourths of respondents relied on evaluation expertise from both inside and outside of the organization. That same year, Fine et al. (2000) found that twenty-nine percent of nonprofits were currently working with a combination of internal and external evaluators. It is unclear why this discrepancy exists; both studies were conducted in the same year and consisted of mail surveys with direct services organizations. Hoefer (2000) surveyed nonprofits in Texas while Fine et al. (2000) surveyed a national sample of nonprofits; perhaps Texas-based nonprofits are particularly likely to work with both internal and external evaluators.

#### *Non-Evaluators*

When the nonprofit does not have a full-time evaluation specialist on staff and does not contract with external evaluation consultants, then evaluation duties often fall to organization leadership such as the executive director, board of directors, or a program

manager. Carman (2007) found that the executive or management staff are responsible for conducting the evaluation in eighty-nine percent of the community development organizations surveyed, seventy-six percent of the social services organization surveyed, and seventy-four percent of the developmental disabilities organizations surveyed. In a similar study by Carman and Fredericks (2008), eighty percent of nonprofits stated that internal executive or management staff were responsible for gathering evaluation data (p. 58). The Board of Directors is sometimes tasked with evaluation duties—twelve percent of nonprofits in the Carman and Fredericks (2008) survey and between two and six percent of community development organizations, social services organizations, developmental disabilities organizations in the Carman (2007) survey. Reed and Morariu (2010) reported that evaluation is the responsibility of the organization's leadership or board in more than half of the organizations surveyed.

In a few cases, volunteers or other staff members are in charge of the evaluation; ten percent of developmental disabilities organizations surveyed relied on volunteers and four percent of social services organizations used volunteers to conduct evaluation (Carman, 2007). It is unclear whether or not these volunteers have prior experience with evaluation. In any case, I suspect it could be difficult for nonprofits to work with volunteer evaluators. Volunteers, by definition, are not paid, and they may not be bound by contracts, which could present challenges when nonprofits seek to keep evaluation projects on tight schedules or if they desire a depth or breadth to the methodology that cannot be reasonably be completed within a volunteer's timeframe.

### **Using the Findings of the Evaluation**

### *How Results are Disseminated*

The evaluation's findings can be disseminated to both internal and external audiences in a number of ways. Morariu et al. (2012) found that all of the organizations surveyed communicated their findings in some way. Most of the organizations reported that they shared evaluation results in updates/reports to the Board of Directors (ninety-one of survey respondents). Additional dissemination formats included: to report to funder(s) on grants and/or contracts (seventy-seven percent of survey respondents), in staff meetings (seventy-six percent of survey respondents), in proposals to funders (seventy-four percent of survey respondents), in an annual report (fifty-one percent of survey respondents), in publications and presentations (forty-seven percent of survey respondents), on the organization's website (forty-one percent of survey respondents), through the organization's social media (thirty-eight percent of survey respondents), in an evaluation report (thirty-three percent of survey respondents), and in the wider media (newspapers, television, radio, etc.; twenty-one percent of survey respondents). An earlier study found that nonprofits felt they had limited capacity to communicate their evaluation results. When asked whether there was "uncertainty about how to gear the presentation of outcome results to different audiences (i.e., clients, funders, general public, board, volunteers, staff)," forty-five percent of nonprofits somewhat agreed or strongly agreed with the statement (United Way of America, 2000, p. 10).

### *How Results are Used*

Morariu et al. (2012) found that all of the organizations used their evaluation results in some manner. Across multiple studies, researchers have found that one of the

most common uses is to improve or adjust existing programs or to plan new programs (e.g., Alaimo, 2008; Carman & Fredericks, 2008; Carman & Millesen, 2004; Fine et al., 2000; Hoefler, 2000; Morariu et al., 2012). These findings are positive because they suggest that organizations are viewing their evaluation activities in “proactive terms” (Carman & Millesen, 2004, p. 15). Another proactive use of evaluation is to inform strategic planning, which has been done by more than two-thirds of nonprofits (Carman & Fredericks, 2008; Carman & Millesen, 2004).

In addition to using findings for program improvement, nonprofits also use evaluation for resource allocation decisions. The United Way of America (2000) found that 74 percent of United Way grantees somewhat agreed or strongly agreed that evaluation should be used in making funding decisions. However, fewer nonprofits actually use evaluation for decisions about funding or budgeting. In 2000, Hoefler found that 16 percent of nonprofits had used evaluation results to transfer funding to a successful program; that same year, Fine et al. found that thirty-six percent of nonprofits used evaluation to make resource allocation decisions (p. 336). By 2012, fifty-eight percent of nonprofits reported using evaluation to make resource allocation decisions (Morariu et al., 2012).

Evaluation can also inform decisions about staffing. In the national survey by Fine et al. (2000), thirty-six percent of nonprofits used evaluation to train or reassign staff. This is a growing trend; twelve years later, Morariu et al. (2012) found that fifty-three percent of nonprofits used evaluation for staffing decisions. However, in 2008, Alaimo’s interviews revealed that only eleven percent of nonprofits use evaluation for

staffing decisions and Carman and Fredericks' (2008) surveys in New York and Ohio found that only eight percent of nonprofits use evaluation for staffing decisions.

Another common use of evaluation results is for reporting and accountability, both for internal and external audiences. For example, between eighty-two percent and eighty-nine percent of survey respondents use evaluation results to report to their Board of Directors (Carman & Fredericks, 2008; Carman & Millesen, 2004; Morariu et al., 2012).

Evaluation results are frequently used for fundraising and marketing purposes, that is, by using results to advocate for funding in proposals to new and existing donors (Alaimo, 2008; Carman & Fredericks, 2008; Carman & Millesen, 2004; Fine et al., 2000; Hoefler, 2000; Morariu et al., 2012). Using results for fundraising is not necessarily problematic unless this becomes a nonprofit's primary impetus for working with an evaluator. In my past experience working as both an internal and external evaluator, nonprofits that are seeking to develop marketing materials will often "fudge the numbers" by inflating positive results and omitting negative results—thus undermining the evaluation's potential to be used to make improvements to the program or organization.

### **Funding for Evaluation**

Evaluation can involve a number of monetary and non-monetary costs, such as an internal evaluator's salary, contracts for external consultants, fees to purchase standardized instruments, software license fees, travel costs, printing fees for publications, and staff time, among others. How do nonprofits pay for evaluation? Nonprofits find funding through a variety of sources, such as grants or from their own

internal operating funds. In a 2003 survey of national nonprofits and accreditation agencies, only thirty percent of respondents said they provided funding to their affiliates to pay for outcome measurement (United Way of America, 2003, p. viii). Funding is sometimes included in grants for the program; twenty-nine percent of the Dallas nonprofits surveyed by Hoefler (2000) said they funded their evaluation that way. However, only eight percent of the Indiana nonprofits surveyed by Carman and Fredericks (2008) received evaluation funding through program grants. A small percentage of nonprofits are able to receive separate grants to fund the evaluation. For example, thirteen percent of the nonprofits in Dallas used a separate grant to provide funds for the evaluation (Hoefler, 2000). In contrast, however, none of the nonprofits in Indiana surveyed by Carman and Fredericks (2008) received separate evaluation grants. When funders do assist with evaluation costs, the funding is most likely to come from foundations and philanthropy (rather than from government agencies or other donors). Nearly one-third (thirty-two percent) of the respondents in the survey by Morariu et al. (2012) reported receiving support from foundations and philanthropy to cover evaluation costs.

Unfortunately, approximately one-third of organizations that are engaging in evaluation receive no financial support from their funders for the evaluation. In 2010, thirty-six percent of nonprofits reported that none of their donors supported evaluation (Reed & Morariu, 2010) and in 2012, thirty-eight percent of nonprofits reported that none of their donors supported evaluation (Morariu et al., 2012). Instead, for many nonprofits, evaluation costs are pulled directly from their own internal operating budget. Nearly two-

thirds (sixty-six percent) of organizations in Indiana relied on internal operating funds for evaluation (Carman & Fredericks, 2008, p. 58) and Hoefler (2000) found that approximately half of the organizations in Dallas funded their evaluation through internal funds (p. 171). When asked how much of their organization's annual budget was spent on evaluation, two-thirds of nonprofits (sixty-six percent) reported spending less than five percent on evaluation. The remaining respondents spent five to ten percent of their budget on evaluation (fifteen percent of nonprofits) or ten percent or more of their budget on evaluation (twelve of nonprofits). Seven percent of nonprofits reported spending none of their budget on evaluation (Morariu et al., 2012, p. 5).

Surprisingly, a relatively high proportion of nonprofits reported that their evaluation does not involve any monetary costs. For instance, Hoefler (2000) found that approximately one-fourth of the nonprofits ascribed no costs to the evaluation (p. 171) and Carman and Fredericks (2008) found that twenty-nine percent of nonprofits reported there were no costs associated with their evaluation activities. These numbers are slightly higher in social services organizations, where thirty-eight percent of those organizations report that their evaluation does not involve any costs (compared to twenty-seven percent of community development organizations and eleven percent of developmental disabilities organizations surveyed; Carman, 2007). Perhaps these organizations are relying on internal staff, such as the Executive Director or Board of Directors, to lead their evaluation, and do not hire professional evaluators. Another related explanation is that these nonprofits may have relatively simple and affordable data collection systems,



such as collecting an annual consumer satisfaction survey, in which the only cost is time from an internal staff member.

### **Barriers to Evaluation**

Multiple studies have asked nonprofits to describe their barriers to conducting evaluation. Across all studies, the top barriers are time, staffing limitations, and money.

#### *Limited Staff Time*

Limited staff time is often described as the largest barrier to participating in and using evaluation. Morariu et al. (2012) found that seventy-one percent of nonprofits felt limited staff time was either a significant challenge or a minor challenge (the remaining 29 percent felt that limited staff time was not a challenge). In their survey of United Way grantees, the United Way of America found that fewer than half of nonprofits (forty percent) somewhat agreed or strongly agreed that there was sufficient staff time available for outcome measurement. When Hoefer (2000) asked nonprofits why evaluations are *not* conducted, forty-eight percent said there was not enough staff time available to conduct an evaluation (p. 171). Limited time is also a barrier when developing an outcome measurement system: fifty percent of United Way grantees somewhat agreed or strongly agreed that there was adequate time to develop and adjust their measurement system (United Way of America, 2000, p. 7).

#### *Limited Staff Expertise*

Most nonprofits find limited staff expertise to be a barrier to their evaluation efforts: thirty-nine percent said limited staff expertise was a significant challenge and an additional forty-three percent said it was a minor challenge (Morariu et al., 2012). When

asked to describe their own evaluation capacity, twenty percent of nonprofits nationwide reported they had “low” evaluation capacity, defined as having a “hard time doing things like tracking program data,” and fifty-one percent reported “moderate” capacity (“for example, we do a good amount of evaluation work in-house, but sometimes we need help;” Morariu et al., 2012, p. 8). Fewer than one-third of nonprofits (twenty-nine percent) reported “high” evaluation capacity, which the survey defined as “regularly complet[ing] entire evaluations in-house” (Morariu et al., 2012, p. 8).

What types of evaluation activities do nonprofits feel least equipped to take on? In 2000, United Way grantees reported having difficulty identifying manageable data collection methods (sixty-eight percent of respondents somewhat agreed or strongly agreed), identifying relevant outcome indicators (sixty-six percent of respondents), identifying appropriate outcomes (sixty-five percent of respondents), and finding/developing appropriate data collection tools (fifty-seven percent of respondents; United Way of America, 2000). An additional forty-eight percent of respondents noted that there were uncertain about how to identify program strengths and weaknesses (United Way of America, 2000). Carman’s 2007 study found that program executives also struggled with unduplicated counts of participants, that is, the organizations counted the number of people served by each program rather than by the organization as a whole. They also “struggled with making decisions about what constitutes 'a person served,' asking if this question meant the number of people referred to the program or the number of people who completed the program” (Carman, 2007, p. 67).

Nonprofits tend to recognize that they need additional education and skills related to evaluation. For example, in the study by Carman and Millesen (2004), more than 40 percent of nonprofits needed “more education about basic evaluation models and concepts as well as better technology to support evaluation activities” (p. 16). More than half of the survey respondents “needed assistance to design and maintain an evaluation system” (Carman & Millesen, 2004, p. 16). Education and training can come from a variety of sources—from internal expertise and prior experience with evaluation; from trainings, webinars, and handbooks provided by funders; or from external evaluation consultants. The United Way of America found that seventy-five percent of its grantees somewhat agreed or strongly agreed that “there was sufficient access to training on outcome measurement” and fifty-nine percent somewhat agreed or strongly agreed that “there was sufficient access to the special expertise needed to resolve specific outcome measurement problems” (United Way of America, 2000, p. 7). However, accessing training from evaluators can be a barrier in itself: forty-four percent of nonprofits felt that “knowing where or how to find an external evaluator” was either a significant or a minor challenge (Morariu et al., 2012, p. 12). Finding an external evaluator is particularly challenging for smaller nonprofits; twenty-one percent said finding an external evaluator was a significant challenge compared to fifteen percent of nonprofits overall (Morariu et al., 2012, p. 12).

#### *Lack of Support from Staff*

Prior research shows that, in some cases, lack of support from the nonprofit’s chief executive officer, chief operating officer, and/or mid-level or frontline staff is a

barrier to engaging in evaluation. For example, fifteen percent of nonprofits nationwide reported that “insufficient support from organization leadership” was a significant challenge; twenty-three percent felt it was a minor challenge (Morariu et al., 2012, p. 12). Carman and Millesen (2004) found that “few encounter staff resistance to data collection activities” (twenty-four percent of nonprofits; p. 16). In 2000, eighty-seven percent of United Way grantees somewhat agreed or strongly agreed that “there was ongoing support by the board of directors” for outcome measurement but eighty-seven percent also somewhat agreed or strongly agreed that “there was ongoing top executive involvement” in outcome measurement (United Way of America, 2000, p. 9).

In addition to reporting a lack of support, a few nonprofits have described open resistance to evaluation. In 2000, thirteen percent of United Way grantees somewhat agreed or strongly agreed that outcome measurement has “lowered the morale of program staff”—although the same study reported that eighty-seven percent of nonprofits somewhat agreed or strongly agreed that “there was ongoing staff commitment to using the results of program outcome measurement” (United Way of America, 2000, p. 6). In 2012, “not having staff who believe in the importance of evaluation” was described as a significant challenge for twelve percent of nonprofits and as a minor challenge for an additional thirty-two percent of nonprofits (Morariu et al., 2012, p. 12). Large organizations “were more likely to identify staff resistance to data collection as a challenge” to evaluation (Carman & Fredericks, 2009, p. 90), perhaps because the internal structures needed to implement data collection strategies can become increasingly complex with more staff members, programs, and locations.

### *Navigating Funder Requirements*

Funders, including government entities and foundations, are exceedingly influential on nonprofits' evaluation practices. When asked why they did not perform an evaluation, for instance, forty-three percent of nonprofits said it was because "their funder did not require one" (Hoefler, 2000, p. 171). When funders do require evaluation, however, interpreting and complying with reporting requirements can be an obstacle to evaluation. More than half (fifty-five percent) of United Way grantees somewhat agreed or strongly agreed that "there were conflicting evaluation methodologies required by funders" (United Way of America, 2000, p. 8). An additional complication is that funders require evaluation, yet "funders tend to ignore evaluation results in their decision making" (twenty-one percent strongly agreed or agreed; Morariu et al., 2012, p. 13). "Funders asking you to report on the wrong data" was named as a significant challenge by twelve percent of nonprofits and as a minor challenge by thirty-two percent of nonprofits (Morariu et al., 2012, p. 12). One way to address these challenges is through conversations with funders about evaluation requirements; on a positive note, four in five nonprofits (eighty-two percent) feel that discussing evaluation with their funders is a worthwhile activity (Morariu et al., 2012).

### *Software*

Today nonprofits can purchase a variety of databases and computer systems to help them enter, store, and analyze findings. Popular software options include Efforts-to-Outcomes, Salesforce, and Wild Apricot, among others. Software can help or hinder a nonprofit's evaluation efforts. In the United Way's (2000) survey, for example, forty-two

percent of nonprofits somewhat agreed or strongly agreed that "there was sufficient computer hardware and software capacity/capability to store and manipulate outcome measurement" (United Way of America, 2000, p. 7)—suggesting that the rest of the respondents felt there was insufficient software to store and manipulate data.

### *Additional Challenges to Evaluation*

As data collection activities have become more detailed and time-consuming, a small percentage of nonprofits feel that “data collection interferes with our relationship with clients” (eighteen agreed or strongly agreed; Morariu et al., 2012, p. 13). A similar percentage of nonprofits—thirteen percent—somewhat agreed or strongly agreed that evaluation has harmed relationships between program staff and clients (United Way of America, 2000).

There are often competing priorities for the organization’s time and financial resources. Morariu et al. (2012) gave survey respondents a list of ten organizational tasks, such as fundraising, communications, human resources, information technology, and staff development, among others, and asked nonprofits to rate their organization’s internal priorities. Fundraising, financial management, and communications were the top three priorities while evaluation and research were the lowest priorities (Morariu et al., 2012).

Additional barriers and disadvantages of evaluation have included: overloading record-keeping capacity, diverting resources from existing activities, leading to a focus on measurable outcomes at the expense of other important outcomes, creating competition for funding with other agencies, and harming relationships with other programs within the nonprofit (United Way of America, 2000). Although we know that

most nonprofits are engaging in evaluation, these findings about the barriers to evaluation demonstrate that most nonprofits are struggling with evaluation.

### **Benefits of Evaluation**

Evaluation can serve many purposes: it can support program and strategic planning, help staff communicate their goals and progress, and serve as a basis for ongoing learning to make the organization's work stronger and more effective. Despite the barriers encountered by some nonprofits, nonprofits have also articulated a number of ways that they have benefitted from evaluation. In his interviews with executive directors of nonprofits, Alaimo (2008) found that, of the organizations engaging in evaluation, 100 percent felt their organization benefitted from evaluation. "Having information to demonstrate program effectiveness" was the most important benefit (mentioned by twenty-four percent of executive directors). Furthermore, the executive directors explained that "organizational learning takes place when results are used and stakeholders realize the benefits of program evaluation... [which is] the turning point for stakeholders" (Alaimo, 2008, p. 83). In an earlier survey by United Way of America (2000), respondents indicated the extent to which outcome measurement had helped their nonprofit. Most commonly, eighty-eight percent of respondents somewhat agreed or strongly agreed that measurement helped the organization communicate results to stakeholders such as board members and donors; eighty-eight percent also somewhat agreed or strongly agreed that measurement helped clarify the intended purpose of the program. Other benefits included: identifying effective practices within the program (eighty-four percent of respondents), successfully competing for funding (eighty-three

percent), enhance our record-keeping system (eighty percent), improve the service delivery of the program (seventy-six percent), share effective practices with programs in other agencies (seventy-two percent), share effective practices with programs in other agencies (seventy-two percent), inform program participants about outcome measurement results (seventy percent), assess staff performance (sixty-four percent), identify staff training needs (sixty-four percent), allocate resources within the program and the agency (sixty-one percent), increase program participants' investment in achieving positive outcomes (fifty-five percent), recruit staff to work in the program (forty-two percent), and recruit volunteers to work in the program (forty-one percent). Similarly, eighty-nine percent of nonprofits would recommend to the director of a similar program that he or she consider implementing program outcome measurement (United Way of America, 2000).

### **Nonprofits' Perspectives on Evaluation**

Up until this point, I have described *how* nonprofits conduct evaluations—what they *do* versus what they *think*. This next section explores how nonprofits think about the broader meaning and purposes behind their evaluation efforts.

#### *Reasons for Engaging in Evaluation*

Nonprofits engage in evaluation for a variety of reasons. When describing their reasons for conducting an evaluation, fifty-seven percent of organizations conducted an evaluation to “ensure control,” forty-two percent of organizations wanted “to use the results for commercial reasons,” and thirty-nine percent of organizations conducted an evaluation “because funders want one done” (Hoefler, 2000, p. 171). Carman (2011)



found that “more than one-third of the nonprofit organizations in this study... conducted evaluations for monitoring or compliance purposes” (p. 365). Fine et al. (2000) found similar results. When asked why they measure outcomes, nearly half (forty-three percent) of nonprofits said they measure outcomes simply because it is a funding requirement, and twenty percent of these nonprofits specifically mentioned a United Way funding requirement for evaluation (Fine et al., 2000, p. 333).

Organizations can be driven to conduct an evaluation by external and/or internal forces. Carman (2011) described that, “for a majority of these organizations, the motivation to do evaluation originated from outside the organization” (Carman, 2011, p. 367). When the United Way asked their grantees which parties had prompted their measurement efforts, ninety-six percent of grantees named the United Way (United Way of America, 2000). Additional drivers of evaluation included agency professional leadership or staff (sixty-one percent), government funders (twenty-five percent), the Board of Trustees or Board of Directors (twenty-two percent), and private funders (nineteen percent; United Way of America, 2000).

Nonprofits also see evaluation as a tool for continuous learning and improvement. Hoefler (2000) reported that half (fifty-one percent) of nonprofits conducted an evaluation because they "are curious" (p. 171). Learning has played a major role in several studies. For example, seventy-nine percent of nonprofits said that “strengthen[ing] future work” was “very important” and an additional eighteen percent of nonprofits said it was “somewhat important” (Morariu et al., 2012, p. 4). Similar percentages of nonprofits felt evaluation is important to “learn whether original objectives were achieved” (seventy-

three said this was very important and twenty-three percent said it was somewhat important; Morariu et al., 2012, p. 4). Learning about implementation and outcomes is especially valued (e.g., Fine et al., 2000; Morariu et al., 2012).

Using evaluation for learning implies that, as programs are modified and improved, clients are served better and their lives improve. Surprisingly, very few nonprofits explicitly say that they engage in evaluation to serve clients better and very few researchers explicitly measure this in their surveys and interviews. Some researchers, such as Fine et al. (2000), have studied a somewhat related concept—whether nonprofits measure client satisfaction with services. In Fine et al.’s (2000) study, only four percent of nonprofits stated that they primarily conducted an evaluation in order to measure participant satisfaction (p. 333). However, nonprofits do engage in evaluation in order to share results with clients. For example, seventy percent of United Way grantees somewhat agreed or strongly agreed that outcome measurement has helped them “inform program participants about outcome measurement results” and fifty-five percent somewhat agreed or strongly agreed that outcome measurement has helped them “increase program participants’ investment in achieving positive outcomes” (United Way of America, 2000, p. 5). Nonprofits also engage in evaluation in order to share results with their peer agencies or with policymakers in order to improve their field as a whole. For example, eighty-four percent of nonprofits nationwide said strengthening organizational practices in the field was very important or somewhat important and fifty-three percent said strengthening public policy was very important or somewhat important (Morariu et al., 2012, p. 4). Similarly, seventy-two percent of United Way grantees

somewhat agreed or strongly agreed that outcome measurement has helped them “share effective practices with programs in other agencies” (United Way of America, 2000, p. 5).

*Evaluation for the Dual Purposes of Accountability and Learning*

Alaimo (2008) described a tension between the external pull for accountability and the internal push for learning and improvement, a tension that still exists in the nonprofit sector. Two studies have sought to better understand how nonprofits deal with this tension by developing typologies of nonprofits based on their motivation for evaluation (Alaimo, 2008; Carman & Fredericks, 2011). After Alaimo (2008) interviewed executive directors of nonprofits, he divided executive directors into three groups (Type I, II, or III) based on how they viewed the purpose or role of evaluation within their organizations. Leaders in Type I were intrinsically motivated to engage in evaluation and viewed evaluation as a tool to improve services. These executive directors were most likely to use a mixed methods approach to evaluation and data collection. Leaders in Type II expressed only a “sporadic” commitment to evaluation, and tended to use satisfaction survey or other single methods to evaluate their work. Leaders in Type III made no effort to evaluate their programs. Carman and Fredericks (2011) found similar clusters among nonprofits. Nonprofits in Cluster 1, for example, tended to connect their evaluation systems with broader management systems, which sometimes involved connecting evaluation with strategic planning initiatives. For other Cluster 1 nonprofits, “evaluation was simply viewed as ‘good business practices’ or an ‘internal management tool to help make decisions’” (Carman and Fredericks, 2011, p. 95).

In recent years, there has been particular emphasis on using evaluation for continuous improvement and learning. For example, nonprofits have begun to express their interest in building a learning culture, and job advertisements seek evaluators to serve as the Director of Learning (rather than a Director of Evaluation, as in past years). Perhaps the clearest indicator that evaluation is valued as a tool for learning is the research results about nonprofits' audiences for the evaluation—in other words, which group of stakeholders the evaluation is designed to serve. In 2000, Fine, Thayer, and Coghlan asked nonprofits to name the audiences of their evaluations. The most common audiences included funders (sixty-nine percent of nonprofits), program staff (sixty-one percent), board members (fifty-six percent), potential funders (forty-four percent), government agencies (twenty-eight), program participants or clients (twenty-four), and for other nonprofits (nine percent). In 2010, Reed and Morariu found similar results: that funders were named as the primary audience for the evaluation. However, by 2012, a shift may have occurred. For the first time, the executive director and Board of Directors were named as the top audiences for the evaluation, over and above funders (Morariu et al., 2012). Twelve percent of nonprofits even stated that funders were not an audience for their evaluation efforts (Morariu et al., 2012). I suspect this trend will continue in the future as more nonprofits seek to use evaluation to inform their own leaders and make strategic decisions, rather than simply as an exercise to fulfill funders' reporting requirements.

## **Chapter 3. Method**

The current study is a secondary analysis of an existing dataset. First, I provide an overview of the larger study from which the data are drawn. Second, I describe the survey items and analysis techniques used in the present study.

### **The Larger Study**

#### *Data Source*

These data come from a larger survey research project, the State of Evaluation 2012 study by Morariu et al. (2012).

#### *Measure*

The full survey included thirty-eight questions about evaluation practices and capacity (twenty-five closed-ended questions and five open-ended questions). The survey also included eight demographic questions (e.g., the number of staff members and the organization's annual operating budget). Specific items used in the current study are described in the following pages. The full survey instrument is provided in the appendix.

#### *Sampling*

When conducting the State of Evaluation 2012 research, Morariu et al. (2012) contacted GuideStar ([www.GuideStar.org](http://www.GuideStar.org)) to purchase contact information for U.S.-based 501(c)3 public charities. The researchers were interested in organizations that had

updated their IRS Form 990 in 2010 or more recently and had provided a contact name and email address. A total of 38,789 organizations met the criteria.

*Survey Administration*

In August 2012, Morariu et al. (2012) sent an initial email invitation to a random sample of 20,000 organizations, followed by three additional reminder emails to representatives who had not yet responded. The survey was open for twenty-five days.

The line chart below displays the number of responses received after each of the reminder messages.

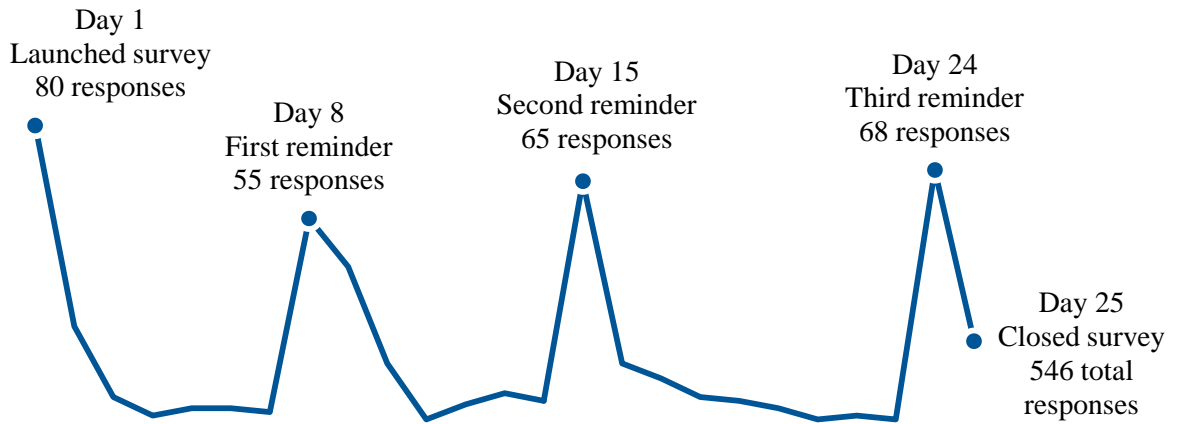


Figure 7. Number of survey responses

The researchers were not able to reach all 20,000 organizations due to a number of logistical and technical reasons. For example several thousand surveys could not be delivered because the receiving email server was busy and not receiving messages; because the receiving email inbox was full; because the receiving organization's firewall blocked the message as spam; or because the email address was incorrect due to a typo or other error. In several cases, Morariu et al. (2012) received automated messages that indicated that the organizations were no longer open. Out of 20,000 organizations, 13,070 organizations received the survey and 546 organizations responses to the survey, for a response rate of four percent.

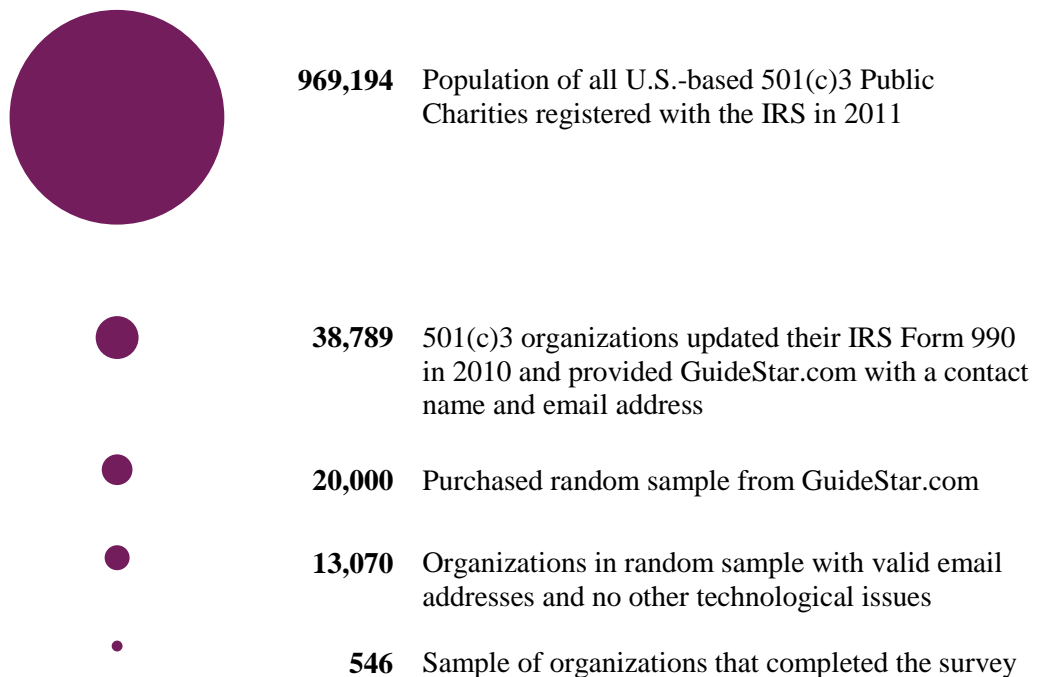
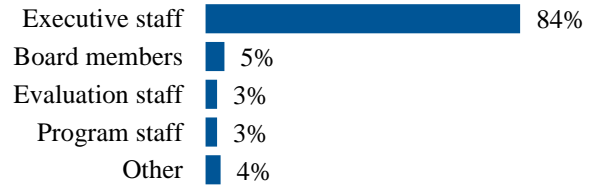


Figure 8. The survey's population, sampling frame, and sample

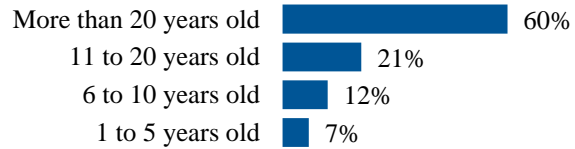
*Participants*

Figure 9 displays demographic characteristics of the survey respondents and their nonprofit organizations.

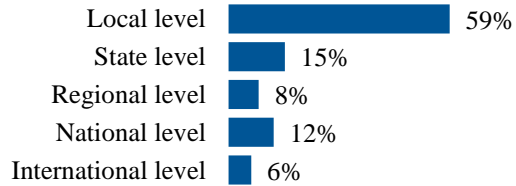
**Respondent's Role**  
Eighty-four percent of survey respondents were executive staff.



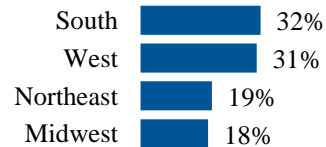
**Organization Age**  
Sixty percent of the nonprofits were more than twenty years old.



**Location of Services Provided**  
Fifty-nine percent of nonprofits provided services locally.



**Location of Headquarters**  
Nonprofits were located throughout the United States.



**Annual Budget Size**  
One-third of the nonprofits were small (annual budgets less than \$500,000).

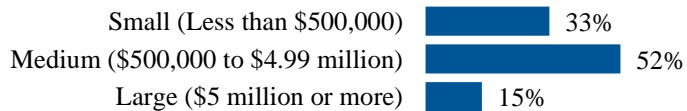


Figure 9. Demographic characteristics of organizations represented in survey



### *Representativeness of the Survey Sample*

To determine the extent to which the survey sample was representative of all U.S.-based 501(c)3 nonprofit organizations, researchers further examined the budget size and geographic distribution of survey respondents.

First, Morariu et al. (2012) compared the organizational budget size of survey respondents to the national average using data from Urban Institute’s National Center for Charitable Statistics. The survey respondents were from larger organizations than is the national average. For example, three-quarters (seventy-five percent) of nonprofits in the Nonprofit Sector in Brief 2011 sample had annual budgets of \$499,999 or less. In comparison, two-thirds (sixty-seven percent) of the nonprofits in the State of Evaluation 2012 sample had annual budgets of \$500,000 or more. The figure below comes from page 16 of the State of Evaluation 2012 report.

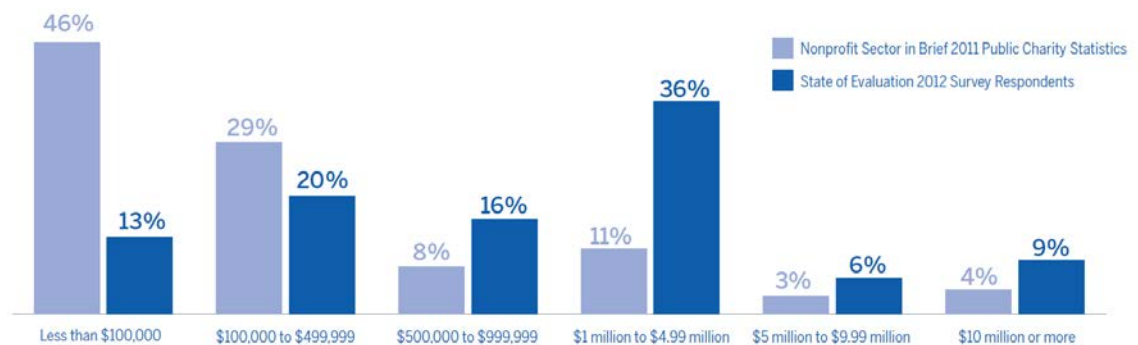


Figure 10. Budget sizes of organizations in sample versus population

Next, I compared the regional distributions of survey respondents to national patterns using data from Urban Institute’s National Center for Charitable Statistics. As shown in the bar charts below, the sample generally mirrors the nationwide distribution of all 501(c)3 nonprofit organizations. For example, thirty-four percent of all nonprofits are located in the south, and thirty-two percent of the survey respondents were located in the south. Midwestern states were slightly underrepresented in the sample (eighteen percent of organizations in the survey sample compared to twenty-four percent of nonprofits nationwide) and western states were slightly overrepresented (thirty percent of organizations in the survey sample compared to twenty-two percent of nonprofits nationwide).

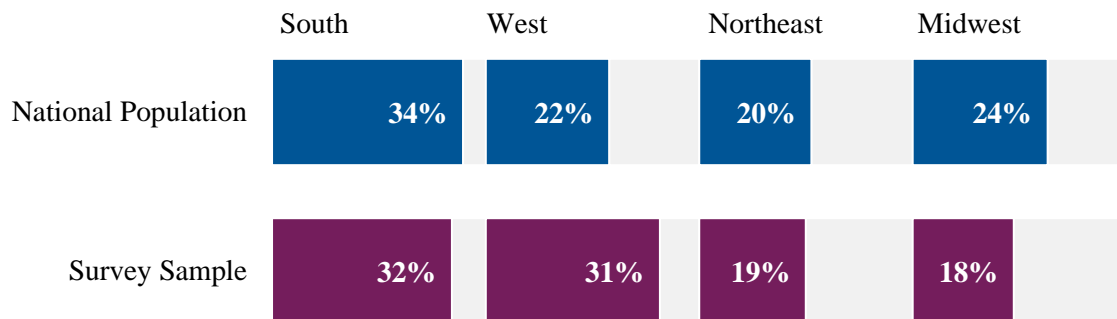


Figure 11. Regional distributions of organizations in sample versus population

I also examined how many states were represented in the study. Nearly every state was represented; the 546 survey responses represented nonprofits from forty-seven states and the District of Columbia. Alaska, Delaware, and Wyoming were the only states not represented in the survey.

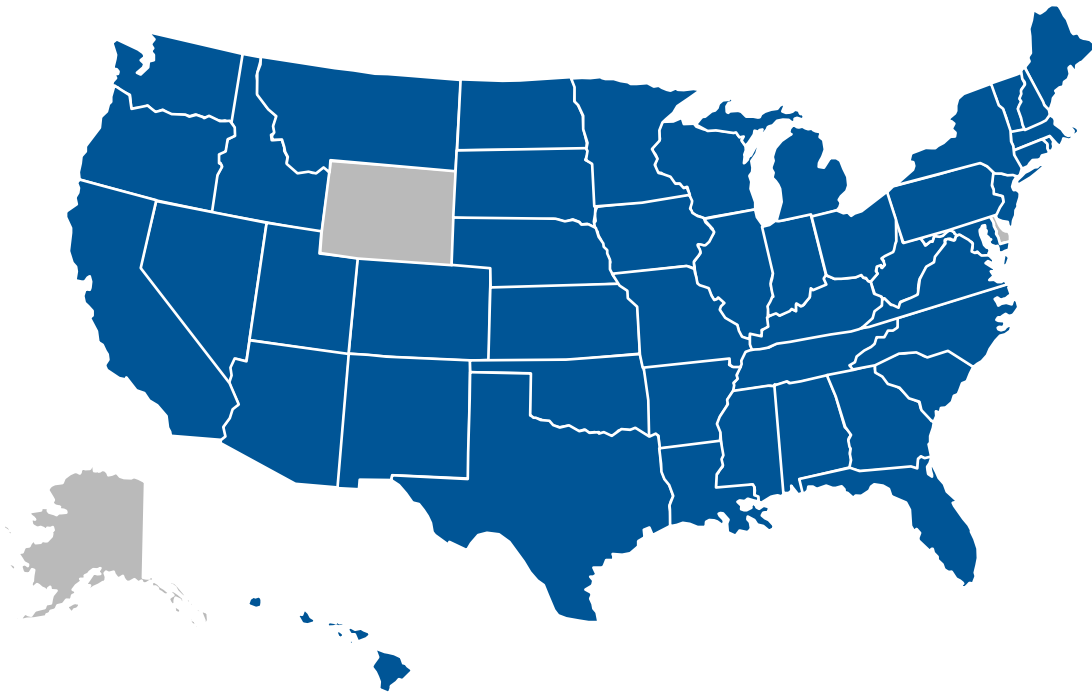


Figure 12. States represented in sample

## The Current Study

The current study is a deeper analysis of two survey items, Question 17 and Question 18/19.

The survey asked nonprofits to describe their experiences working with an external evaluator through a two-part question: first, a closed-ended question about their satisfaction working with the external evaluator; and second, an open-ended question asking nonprofits to elaborate on their experience. The skip pattern is illustrated in the following figure.

17. If your organization worked with an external evaluator in 2011, what was your experience?

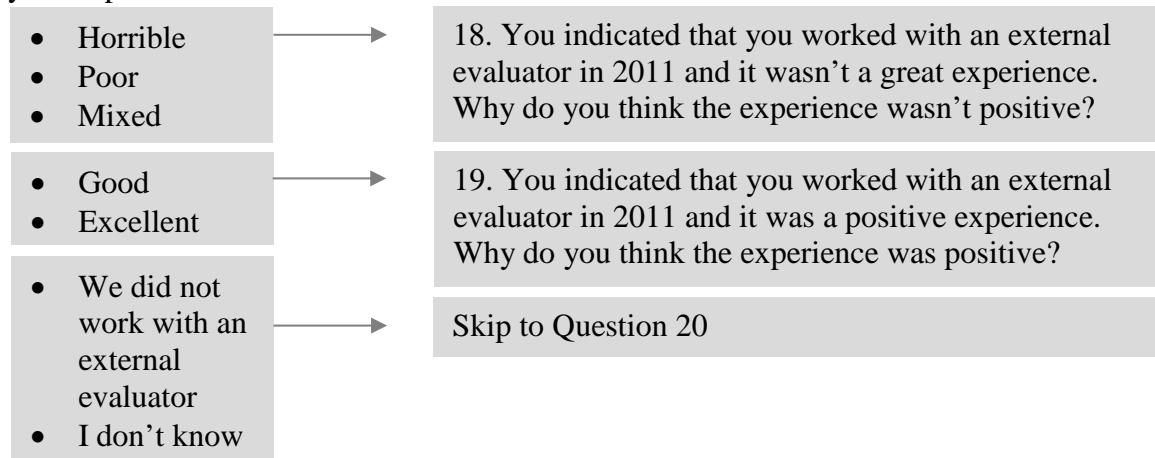


Figure 13. Skip pattern for two-part survey question

Anfara, Brown, and Mangione (2002) advocate for “publicly disclosing decisions made during the research process” (p. 30), so I have documented my decision making processes in the following sections.

**Research Question 1: How many nonprofits are working with internal and/or external evaluators?**

To answer this question, I tabulated the percentage of nonprofits working with internal and/or external evaluators.

**Research Question 2: How do nonprofits describe their experiences with external evaluators?**

*Step 2a. Examine closed-ended ratings of satisfaction with external evaluator*

First, I examined the closed-ended survey item (Question 17). The survey asked nonprofits to rate their experience working with an external evaluator as *excellent*, *good*, *mixed*, *poor*, or *horrible*. I tabulated the percentage of nonprofits that selected each response option.

*Step 2b: Begin coding process by skimming the open-ended responses*

Next, I examined the corresponding open-ended survey questions. My goal was to code the open-ended responses and organize codes into a hierarchy.

*Step 2c: Calculate descriptive statistics for open-ended responses*

After skimming the responses, I calculated simple metrics on each of the open-ended questions, such as the number of responses to each question and the typical length of responses.

*Step 2d: Begin coding negative responses*

To better understand how nonprofit survey respondents described their experiences working with external evaluation consultants, I coded the survey responses. According to Saldana's (2013) *Coding Manual for Qualitative Researchers*, "a code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data" (p. 3). Glesne (2011) describes how "line-by-line coding helps to immerse [the researcher] in the data and discover what concepts they have to offer" (p. 195). I began line-by-line coding of the negative responses first because there were fewer responses (forty negative responses compared to ninety-one positive responses).

When coding the survey responses, I assigned both *emic* and *etic* codes. Emic codes are based on how the research participant makes sense of his or her own experiences (Maxwell, 2005). One example is an *in vivo* code in which the researcher constructs a code from the participant's own language (R. Sheridan, personal communication, November 11, 2011). Etic codes represent the researcher's interpretation of what is going on (Maxwell, 2005).

*Step 2e: Begin coding positive responses*

After coding the negative responses, I began tagging the positive responses (the open-ended responses from Question 18).

*Step 2f: Revisit and refine the codes*

I followed Glesne's (2011) instructions to "try out your new coding scheme... and see how it works... What new codes are added?" (p. 195). I spent several sessions reading, refining, and adding nuances to the coding structure. "Go back and forth... until you are no longer adding substantially more codes," advises Glesne (2011, p. 195).

*Step 2g: Build code hierarchy*

During this step of the analysis, I organized the codes into a hierarchy. Glesne (2011) writes that "arranging your codes in hierarchies is part of the analysis process" (p. 195).

## Chapter 4. Results

The purpose of this study is two-fold: 1) to understand how many nonprofits are working with internal and/or external evaluators and 2) to learn how they describe their experiences working with evaluators. Chapter 4 presents results from the study.

### **Research Question 1: How many nonprofits are working with internal and/or external evaluators?**

As shown in the venn diagram, nine percent of nonprofits reported only working with their internal evaluator; eighteen percent worked only with an external evaluator; and seven percent worked with both internal and external evaluators. Taken together, sixteen percent of nonprofits worked with internal evaluators in some way and twenty-five percent worked with external evaluators in some way.



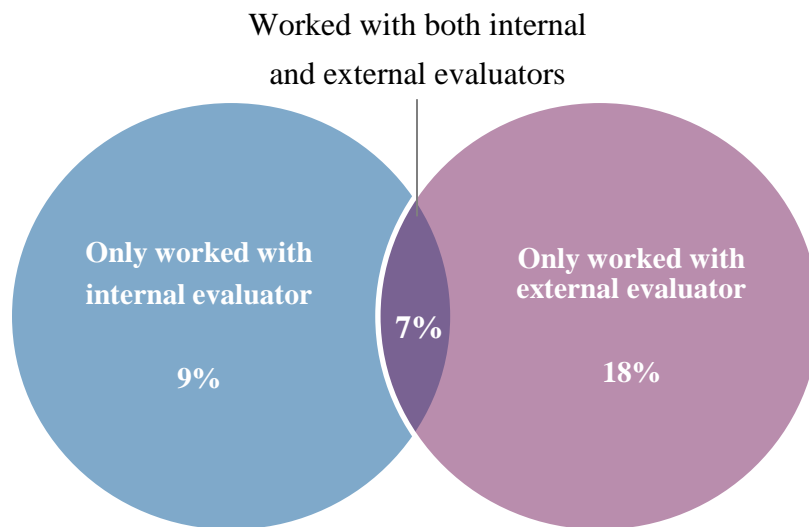


Figure 14. Percentage of nonprofits working with internal and/or external evaluators

**Research Question 2: How do nonprofits describe their experiences with external evaluators?**

*Step 2a. Examine closed-ended ratings of satisfaction with external evaluator*

As shown in the stacked bar chart, more than two-thirds felt their experience with an external evaluator was positive: sixty-nine percent said the experience was either *excellent* or *good*. To add context, I added sample quotes from the corresponding open-ended question that begin to explain why the experience was or was not positive.

**If your organization worked with an external evaluator in the past year, what was your experience?**

**Why do you think the experience was/wasn't positive?**



Figure 15. Nonprofit ratings of experiences with external evaluators

*Step 2b: Begin coding process by skimming the open-ended responses*

After skimming the open-ended survey responses, a few responses stuck out to me, such as “The culture of the organization is more collaborative than the evaluator,” “Difficult to get a hold of because of his many other commitments; tardy evaluation report,” and “They did their own thing and didn't know the project or culture of the community.” These examples, in particular, illustrate the importance of adapting to the nonprofit’s unique organizational culture. I systematically recorded these reflections, as shown in the screenshot below.

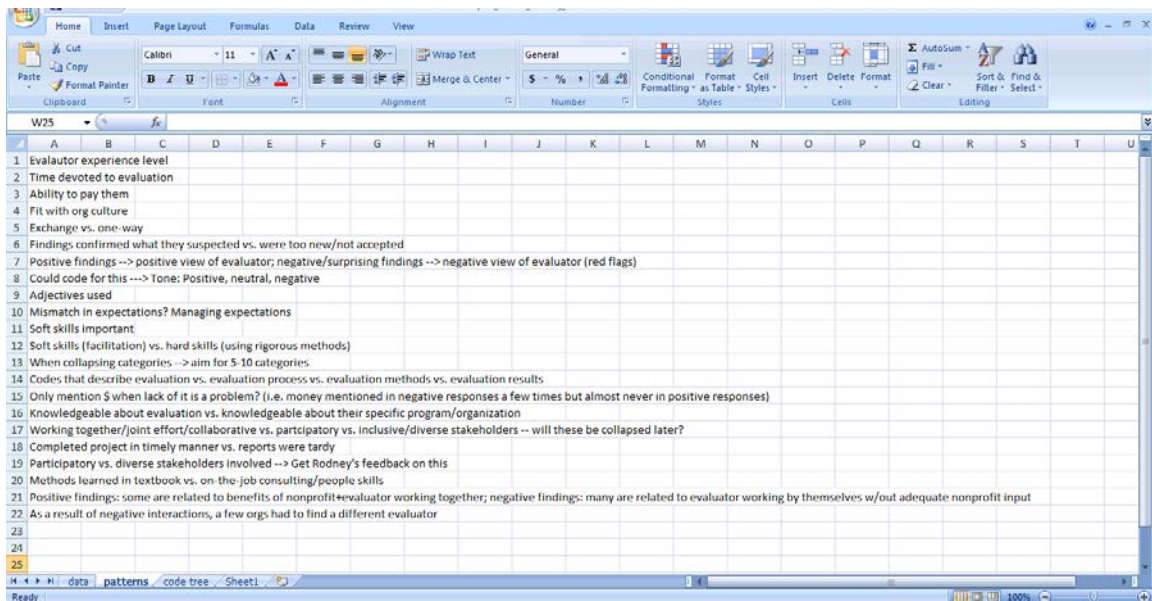


Figure 16. Reflections about emerging findings

*Step 2c: Calculate descriptive statistics for open-ended responses*

There were ninety-one positive responses to Question 18 and forty negative responses to Question 19. Positive responses ranged from two to one hundred thirty-three words in length (median = nineteen words) while negative responses ranged from two to sixty words in length (median = fifteen words). An example of a shorter response was “Competency and professionalism issues” (a negative response with four words). An example of a longer response is “They were professional, took the time to consider our missions statement, needs, demographics, and desired outcomes, educated the staff about their evaluation process and completed the project in a timely manner” (a positive response with thirty words).

The box plot illustrates how the positive responses were slightly longer than the negative responses.

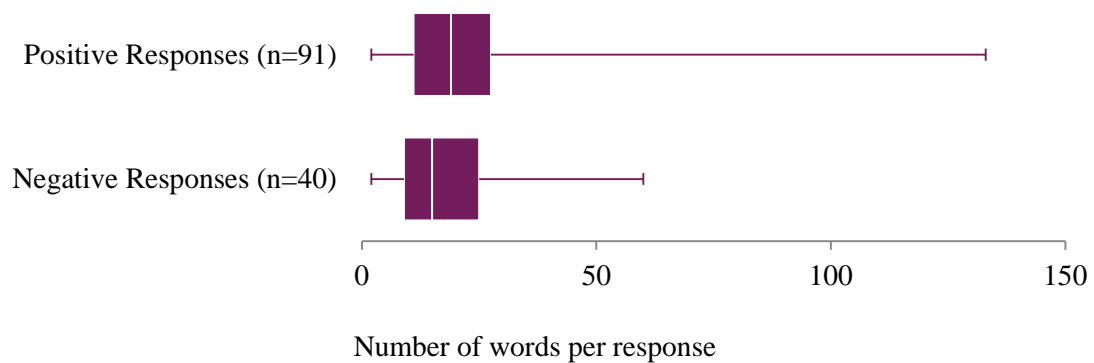


Figure 17. Word counts for open-ended responses

Step 2d: Begin coding negative responses

Most of the nonprofits offered multiple explanations as to why their experience with an external evaluator was unsatisfying. Two examples are shown below.

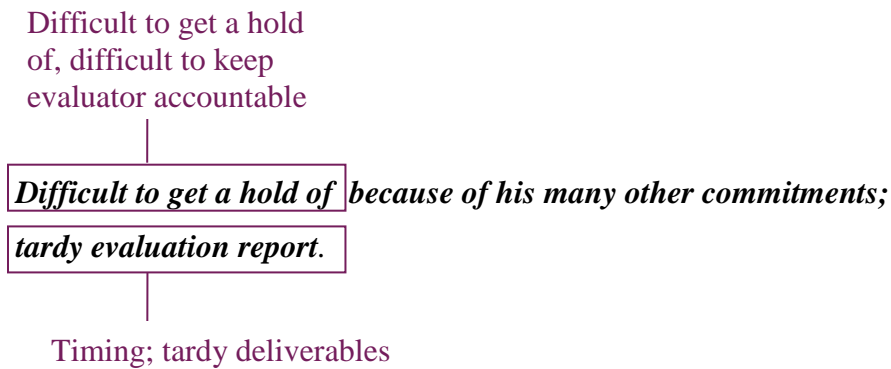
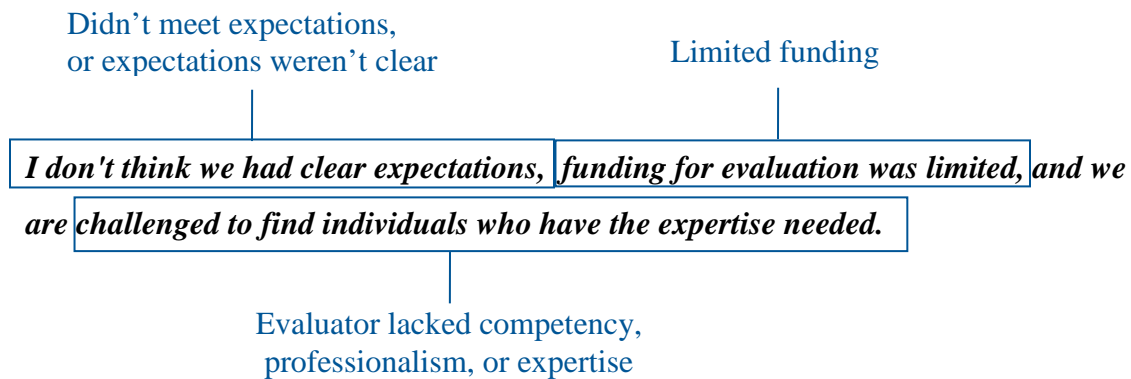


Figure 18. Multiple explanations as to why evaluation experiences were negative

*Step 2e: Begin coding positive responses*

Again, I found that most nonprofits offered multiple reasons for their positive experience. An example is shown below.

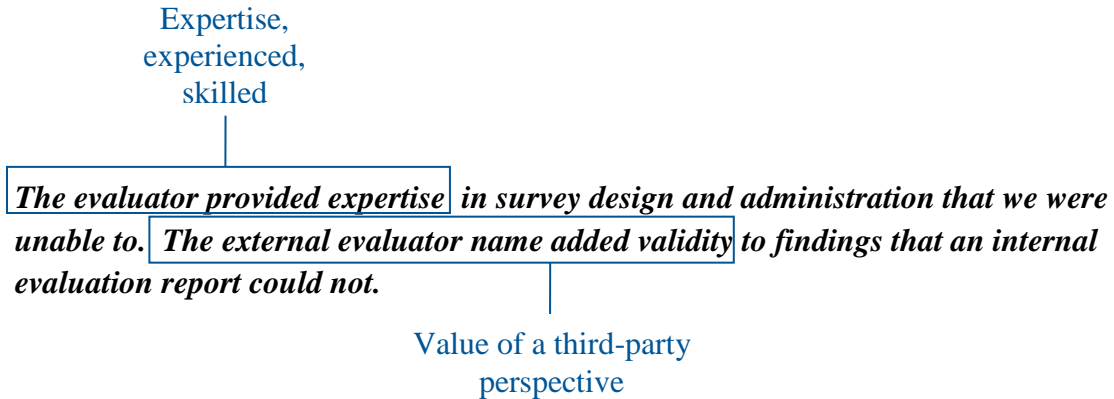


Figure 19. Multiple explanations as to why evaluation experiences were positive

I noticed connections between the positive and negative responses in which responses seemed to be exact opposites of each other. One example is displayed below.

Table 1. Corresponding positive and negative experiences

Positive Response:	Negative Response:
<b><i>We received a detailed report that answered questions we set forth in the plan for evaluation.</i></b>	<b><i>We never got a full report and they kept insisting we needed to pay more money to get what we needed.</i></b>

*Step 2f: Revisit and refine the codes*

I read, re-read, and re-coded the open-ended responses several times. I categorized the open-ended survey responses into twenty-two distinct positive codes and fifteen distinct negative codes. Positive codes are listed below, with corresponding negative quotes in parentheses as applicable.

1. The evaluator was experienced, skilled, competent, knowledgeable, professional, seasoned, prepared, or provided expertise. The evaluator brought more experience than was available in-house, which increased the nonprofit's capacity. (Related negative code: The evaluator lacked expertise, competency, or professionalism.)
2. The evaluator possessed background knowledge of the organization, program, field, or content area. (Related negative code: The evaluator was unfamiliar or disconnected from the organization, the organization's culture, or the programmatic area.)
3. Soft skills (e.g., good listener, excellent facilitator, organized, engaged, or committed). (No corresponding negative code.)
4. Value of a third party perspective: New, fresh, neutral, valuable, void of biases, another point of view; Not slighted by internal pressures. (Related negative code: Stressful to open up to an outsider.)
5. The evaluation process was useful, educational, or helpful and contributes to continued success. (No corresponding negative code.)
6. Met expectations; met needs, expectations and goals were clear, they did what we asked them to do, performed the required tasks; met deliverables. (Related

negative code: Didn't meet expectations, or expectations weren't clear up front; perfunctory; unsatisfying experience.)

7. The evaluator guided the nonprofit to set up processes or think through new ideas. (No corresponding negative code.)
8. Appropriate depth and/or breadth. Holistic assessment; thorough. (Related negative code: Insufficient depth and/or breadth: Not as deep as they would have liked; or, too complicated.)
9. “Good” or tailored overall approach. Good approaches to evaluation. Tailored approach in which the evaluator adapts methods to fit nonprofit's unique needs. (Related negative code: Not tailored to nonprofit's unique needs. Inflexible. Evaluator focused on own needs.)
10. Resources: Effective and efficient use of resources, ample resources. (Related negative code: Limited funding. Challenges with pro bono or volunteer evaluator.)
11. Process was useful: The evaluation process was useful, educational, and helpful, and contributed to continued success. (No corresponding negative code.)
12. Findings were useful and used. The nonprofit gained new information or new insights about best practices. (Related negative code: Findings were disseminated or published without the nonprofit's permission.)
13. Excellent and practical recommendations. (No corresponding negative code.)



14. Methods and analysis: The evaluator developed measurable criteria, thoroughly analyzed data, and/or performed a meaningful analysis. (No corresponding negative code.)
15. Excellent and detailed reports. (Related negative code: Never received report.)
16. Adequate time and/or attention was given to the nonprofit or the evaluation process. The evaluator held regular meetings and may have met with the nonprofit whenever necessary. As a result, the evaluator was described as thoughtful, accessible, or responsive. (Related negative code: Insufficient time and/or attention was given. The evaluator may have been difficult to get a hold of. A few nonprofits felt the evaluator did not devote enough time to understand their organization.)
17. Good communication. (Related negative code: Poor communication.)
18. Deliverables were completed in a timely manner. (Related negative code: Deliverables were tardy; the process was slower than desired.)
19. Negative code: The evaluation was labor-intensive for the organization. (No corresponding positive code.)
20. Collaborative. The evaluation process was a joint effort between evaluators and nonprofits. There was close coordination as the evaluator and nonprofits worked together and exchanged ideas. (Related negative code: Not collaborative. The evaluators and nonprofits did not reach consensus on decisions throughout the evaluation planning or implementation. Results may not have been vetted or co-created.)

21. Staff were trained or educated about evaluation. The evaluator's goal was to help the staff become more independent or self-sufficient when working on the evaluation. (No corresponding negative code.)
22. Participatory approach with a diversity of stakeholders. The evaluator brought staff of different levels to the table. (No corresponding negative code.)
23. Red flags: The evaluation confirmed positive findings. Information was used for fundraising/marketing. Doing an evaluation set us apart from other nonprofits in the field. (No corresponding negative code.)

*Step 2g: Build code hierarchy*

After completing this coding process, I organized the child codes into six parent codes. I distinguished codes based on whether they applied to the individual evaluation consultant or the broader evaluation process as a whole. For example, the first parent code (the evaluator's experience, knowledge, and soft skills) relates to the evaluation consultant while the second parent code (qualities of the evaluation process as a whole) relates to the evaluation process. I also grouped codes based on skillsets. For example, the third parent code (the methods, results, reports, and recommendations) relates to methodological skills while the fourth parent code (project management: timing and communication) relates to project management skills. The six parent codes include:

1. The evaluator's experience, knowledge, and soft skills
2. Qualities of the evaluation process as a whole
3. The methods, results, reports, recommendations
4. Project management: Timing and communication

5. Collaboration (or lack of collaboration) with staff
6. “Red Flags:” Evaluation for the purpose of fundraising or reputation rather than for learning and improvement

The final code tree is displayed on the following pages. Each parent code is displayed in a separate table with a fuller description and several examples. Diametrically opposed positive and negative experiences are displayed next to each other, where applicable.

#### *The Evaluator's Experience, Knowledge, and Soft Skills*

The most frequent topic mentioned by nonprofits was the individual evaluator’s prior experience, knowledge of the programmatic area and/or evaluation knowledge, or soft skills. For example, of the ninety-one nonprofits that said their recent experience with evaluation was *excellent* or *good*, one in four (twenty-five percent) ascribed that experience to the consultant’s experience, skills, or competency. Sometimes the evaluator’s experience was especially appreciated because they added more capacity than was available inside the nonprofit, which justifies the cost and time commitment of contracting with an outside consultant. One example is, “She provided expertise that we would have lacked otherwise.”

Similarly, when describing their negative experiences with evaluation, twenty-eight percent of nonprofits also mentioned the particular evaluator’s experience, knowledge, or soft skills—but not through a complimentary light. For example, nonprofits lamented the “competency and professionalism issues” that were present during their evaluation experience. Table 2 provides an overview of these responses.

Table 2. The evaluator's experience, knowledge, and soft skills

Positive Experiences	n (%)	Negative Experiences	n (%)
<p><b>The evaluator was experienced,</b> skilled, competent, knowledgeable, professional, seasoned, prepared, or provided expertise. The evaluator brought more experience than was available in-house, which increased the nonprofit’s capacity.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “She provided expertise that we would have lacked otherwise.”</li> <li>• “They were professional.”</li> <li>• “The evaluator was excellent.”</li> <li>• “They brought a depth of experience in performance management that we did not have in house.”</li> </ul>	<p>23 (25%)</p>	<p><b>The evaluator lacked expertise,</b> competency, or professionalism.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “Competency and professionalism issues.”</li> <li>• “The evaluator seemed to be in over their head.”</li> <li>• “The Certification Surveyor lacked proper training to do Certification Survey.”</li> <li>• “We are challenged to find individuals who have the expertise needed.”</li> </ul>	<p>11 (28%)</p>
<p><b>The evaluator possessed background knowledge</b> of the organization, program, field, or content area.</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “The evaluator's knowledge of our field.”</li> <li>• “Our external evaluators come from a membership organization that we belong to. Those evaluators are and have been involved with the same work that we are doing and understand where we are in our growth.”</li> </ul>	<p>11 (12%)</p>	<p><b>The evaluator was unfamiliar or disconnected</b> from the organization, the organization’s culture, or the programmatic area.</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “The evaluator was far removed from our agency, as she worked from home, and hardly visited the agency to get a real feel of our programs.”</li> <li>• “Hard to communicate our needs and match with</li> </ul>	<p>7 (25%)</p>

Positive Experiences	n (%)	Negative Experiences	n (%)
<ul style="list-style-type: none"> <li>“The evaluators were involved with like Agencies in the past, familiar with mission and vision.”</li> </ul>		their interests/priorities.”	
<p><b>Soft skills</b> (e.g., good listener, excellent facilitator, organized, engaged, or committed).</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>“Excellent facilitator.”</li> <li>“Organized, engaged, and committed to the research they conducted for us and the analysis of their findings.”</li> </ul>	5 (5%)	<i>n/a – No corresponding negative category</i>	

*Qualities of the Evaluation Process as a Whole*

In addition to naming specific qualities of the evaluation consultant, nonprofits also mentioned qualities of the evaluation process as a whole: their opinion of hearing a third party perspective of their organization, whether the evaluation met their expectations, and whether the evaluation reached the appropriate depth and breadth needed by their organization at that point in time, among other categories.

Twenty-one percent of the nonprofits that had positive evaluation experiences named having a third party evaluation perspective as being valuable. They described this outside perspective as being new, fresh, neutral, valuable, void of biases, and not slighted

by internal pressures. For example, one survey respondent wrote, “[External evaluation] brings a new perspective to the environment and the outlook of findings is neutral.”

Another area—mentioned by three percent of nonprofits with positive evaluation experiences and eighteen percent of nonprofits with negative evaluation experiences—was the match (or mismatch) between the overall evaluation approach and the nonprofit’s needs. Those with positive experiences described the approach used as tailored, or simply “good,” because the evaluator adapted methods to fit the nonprofit’s unique needs. For example, one person wrote that the evaluator was “willing to change and edit [the] evaluation tool so that it worked for our programs.”

Eighteen percent of nonprofits with negative evaluation experiences attributed their experience to the evaluation not being tailored to their needs. They described the evaluation process as “inflexible” or said the evaluation focused on his or her own needs rather than the organization’s needs. For example, a respondent wrote that “the process seemed inflexible and did not always fit our reality and programs.” Another respondent wrote, “Sometimes evaluators come in with a theory in mind and work to prove it rather than learning from the study.”

Table 3 illustrates the responses that were associated with the evaluation process holistically (rather than characteristics of the individual consultant working on the evaluation project). For example, one in five nonprofits (twenty-one percent) that had an excellent or good experience with evaluation described how an outside “third party” perspective was valuable.

Table 3. Qualities of the evaluation process as a whole

Positive Experiences	n (%)	Negative Experiences	n (%)
<p><b>Value of a third party perspective.</b> New, fresh, neutral, valuable, void of biases, another point of view. Not slighted by internal pressures.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “External evaluators bring fresh perspectives.”</li> <li>• “It brings a new prospective to the environment and the outlook of findings is neutral.”</li> <li>• “A 3rd party perspective and new information/ideas about best practices.”</li> <li>• “Bringing in an outside perspective gave us an objective view on the performance and successes of the organization.”</li> </ul>	<p>19 (21%)</p>	<p><b>Stressful to open up to an outsider.</b></p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “While we knew we needed a ‘makeover’ and major staff changes, it was stressful to open up to outsiders and admit the need.”</li> </ul>	<p>1 (3%)</p>
<p><b>Met expectations;</b> met needs, expectations and goals were clear, they did what we asked them to do, performed the required tasks; met deliverables.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “Performed as project was scoped.”</li> <li>• “Met all of our goals.”</li> </ul>	<p>7 (11%)</p>	<p><b>Didn't meet expectations,</b> or expectations weren't clear up front; perfunctory; unsatisfying experience.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “I don't think we had clear expectations.”</li> <li>• “It was perfunctory and not satisfying.”</li> </ul>	<p>4 (10%)</p>
<p><b>The evaluator guided the nonprofit</b> to set up processes or think through</p>	<p>7 (8%)</p>	<p><i>n/a – No corresponding negative category</i></p>	

<p>new ideas.</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Evaluator helped us to set up processes and was very helpful in thinking through new ideas.”</li> </ul>	
<p><b>Appropriate depth and/or breadth.</b> 3 Holistic assessment; thorough. (3%)</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “It was a holistic organizational assessment.”</li> <li>• “Having an international accreditation association look at all of your systems, processes, policies, core domains, interview our staff, our students and our clients provided us the objective feedback.”</li> </ul>	<p><b>Insufficient depth and/or breadth:</b> 2 Not as deep as they would have liked; or, too complicated. (5%)</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Systems developed too complicated. We are simplifying and streamlining now with great success.”</li> </ul>
<p><b>“Good” or tailored overall approach</b> in which the evaluator adapts methods to fit nonprofit's unique needs. 3 (3%)</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “Willing to change and edit evaluation tool so that it worked for our programs.”</li> <li>• “Group we worked with... had good approaches to evaluation.”</li> </ul>	<p><b>Not tailored to nonprofit's unique needs.</b> 7 Inflexible. Evaluator focused on own needs. (18%)</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “The process seemed inflexible and did not always fit our reality and programs.”</li> <li>• “Sometimes evaluators come in with a theory in mind and work to prove it rather than learning from the study.”</li> </ul>
<p><b>Effective and efficient use of</b> 2</p>	<p><b>Limited funding.</b> Challenges 8</p>



<b>resources; ample resources.</b> (2%)	with pro bono or volunteer (20%) evaluator.
<p>Examples:</p> <ul style="list-style-type: none"> <li>• “By contracting with external evaluators with relevant expertise with regard to the project, we are able to effectively and efficiently utilize resources and ensure quality outputs.”</li> <li>• “Ample resources were provided to get the job done.”</li> </ul>	<p>Examples:</p> <ul style="list-style-type: none"> <li>• “Evaluator worked for us on a pro-bono basis, so there was an issue about holding this person accountable to our scheduling needs.”</li> <li>• “Backed out because of the amount of work involved as compared to the payment.”</li> </ul>

*The Methods, Results, Reports, and Recommendations*

The previous categories describe skill sets that are often inherent to an individual (e.g., soft skills, which can be difficult to teach) or require hands-on practice (e.g., adapting methods and instruments so that the overall approach to the evaluation is tailored to the nonprofit’s needs). The next category mentions the evaluation’s methods, results, reports, and recommendations—specific deliverables or pieces of the evaluation that can often be learned through classes, workshops, or conference sessions.

Of the nonprofits that had positive evaluation experiences, four percent noted that they received “excellent” or “detailed” evaluation reports. “We received a detailed report that answered questions we set forth in the plan for evaluation,” described one of the survey respondents. Of the nonprofits that had negative experiences with evaluation, three percent said they never received a report. “We never got a full report and they kept

insisting we needed to pay more money to get what we needed,” explained one of these nonprofits.

Evaluation reports often contain a set of recommendations for improving the program or organization. Eight percent of survey respondents named their “excellent” and “practical” recommendations as a reason for their positive experience. One respondent wrote that “the input was invaluable.”

Table 4 shows responses that dealt with the evaluation’s particular methods, results, reports, and/or recommendations.

Table 4. Methods, results, reports, and recommendations

Positive Experiences	n (%)	Negative Experiences	n (%)
<p><b>Process was useful:</b> The evaluation process was useful, educational, and helpful, and contributed to continued success.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “It was a very helpful process.”</li> <li>• “The process was educational and extremely helpful.”</li> </ul>	<p>12 (13%)</p>	<p><i>n/a – No corresponding negative category</i></p>	
<p><b>Findings were useful and used.</b> The nonprofit gained new information or new insights about best practices.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “The findings from this external evaluation were used to improve our processes.”</li> </ul>	<p>9 (10%)</p>	<p><b>Findings were disseminated or published without the nonprofit's permission.</b></p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Data were used without acknowledging our organization in</li> </ul>	<p>1 (3%)</p>

<ul style="list-style-type: none"> <li>• “We have a report from which we have developed a number of committees to make improvements to the building and to the activities conducted. We have made some large steps in improving both areas of the evaluation.”</li> </ul>	<p>publications.”</p>
<p><b>Excellent and practical recommendations.</b></p> <p>7 (8%)</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “Excellent suggestions on how to improve communications outside the organization.”</li> <li>• “The input was invaluable.”</li> <li>• “She was able to provide feedback which enabled us to modify our work.”</li> </ul>	<p><i>n/a – No corresponding negative category</i></p>
<p><b>Methods and analysis:</b> The evaluator developed measurable criteria, thoroughly analyzed data, and/or performed a meaningful analysis.</p> <p>6 (7%)</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “They were very thorough and thoughtful in the evaluation methodology and results.”</li> <li>• “She was... able to help us develop the metrics to provide meaningful analysis of a multi-year pilot program.”</li> <li>• “Defined clear measurable objectives.”</li> </ul>	<p><i>n/a – No corresponding negative category</i></p>

<b>Excellent and detailed reports.</b>	4 (4%)	<b>Never received report.</b>	1 (3%)
Examples:		Example:	
<ul style="list-style-type: none"> <li>• “We received excellent and detailed reports of our performance.”</li> <li>• “We received a detailed report that answered questions we set forth in the plan for evaluation.”</li> <li>• “Very helpful report.”</li> </ul>		<ul style="list-style-type: none"> <li>• “We never got a full report and they kept insisting we needed to pay more money to get what we needed.”</li> </ul>	

*Project Management: Timing and Communication*

Some of the nonprofits mentioned issues of project management as being central to their evaluation experience, such as good (or poor) timeframes and communications about the project. Of the nonprofits that had positive evaluation experiences, one in ten (ten percent) appreciated their external evaluation consultant’s devotion to the project. These consultants were described as thoughtful, accessible, or responsive, and held regular meetings—or met as often as was necessary. For example, one nonprofit said that “the evaluator spent adequate time assessing staff roles and getting to know the true organizational climate.” Another consultant “met with us whenever necessary.” Unfortunately, of the nonprofits that had negative evaluation experiences, one in five (twenty percent) noted that the external evaluation consultant was difficult to get a hold of. Some of these consultants were “difficult to get a hold of because of [his or her] many other commitments.”

A handful of nonprofits mentioned the general communication styles between the evaluation consultant and the organization. Nonprofits that had positive evaluation experiences simply described the communication as being “good” (two percent of respondents); nonprofits that had negative experiences described the communication as being “not always ideal” (five percent of respondents).

Table 5 considers aspects of project management, such as the project’s timeframe and communications.

Table 5. Project management: Timing and communication

Positive Experiences	n (%)	Negative Experiences	n (%)
<p><b>Adequate time and/or attention was given</b> to the nonprofit or the evaluation process. The evaluator held regular meetings and may have met with the nonprofit whenever necessary. As a result, the evaluator was described as thoughtful, accessible, or responsive.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “The evaluator spent adequate time assessing staff roles and getting to know the true organizational climate.”</li> <li>• “She met with us whenever necessary.”</li> </ul>	<p>9 (10%)</p>	<p><b>Insufficient time and/or attention was given.</b> The evaluator may have been difficult to get a hold of. A few nonprofits felt the evaluator did not devote enough time to understand their organization.</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Difficult to get a hold of because of his many other commitments.”</li> <li>• “When the person is not onsite, it’s hard to keep them (and yourself) focused on evaluation.”</li> </ul>	<p>7 (18%)</p>

<p><b>Good communication.</b> 2 (2%)</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Communication was good between the evaluation and our agency.”</li> </ul>	<p><b>Poor communication.</b> 2 (5%)</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Communication between the evaluator and the organization was not always ideal.”</li> </ul>
<p><b>Deliverables were completed in a timely manner.</b> 1 (1%)</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “Completed the project in a timely manner.”</li> </ul>	<p><b>Deliverables were tardy;</b> the process was slower than desired. 4 (10%)</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “Tardy evaluation report.”</li> <li>• “Very slow to get results.”</li> </ul>
<p><i>n/a – No corresponding positive category</i></p>	<p>The evaluation was <b>labor-intensive</b> for the organization. 2 (5%)</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “The documentation they needed was difficult for my staff to obtain.”</li> </ul>

*Collaboration (or Lack of Collaboration) with Nonprofit Staff*

A range of different collaboration styles between evaluators and nonprofit survey respondents took place during the evaluation processes. Of the nonprofits with positive evaluation experiences, twelve percent described the evaluation process as being a joint effort that involved close coordination in which nonprofit and evaluators worked together and shared ideas. “The evaluation was a real collaboration between the researcher and the

staff from start to finish,” described one respondent. Another person said, “The evaluator and the staff worked together to create and implement the evaluation. It was a joint effort.”

Table 6 discusses elements of collaboration, cooperation, and co-creation between the evaluator and the nonprofit’s staff.

Table 6. Collaboration (or lack of collaboration) with staff

Positive Experiences	n (%)	Negative Experiences	n (%)
<p><b>Collaborative.</b> The evaluation process was a joint effort between evaluators and nonprofits. There was close coordination as the evaluator and nonprofits worked together and exchanged ideas.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “Because we work very well with our external evaluators and found it to be collaborative and positive.”</li> <li>• “The evaluator and the staff worked together to create and implement the evaluation. It was a joint effort.”</li> <li>• “The evaluation was a real collaboration between the researcher and the... staff from start to finish.”</li> </ul>	11 (12%)	<p><b>Not collaborative.</b> The evaluators and nonprofits did not reach consensus on decisions throughout the evaluation planning or implementation. Results may not have been vetted or co-created.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “We could not reach into consensus of clear outcome of the work performed.”</li> <li>• “Because the culture of the organization is more collaborative than the evaluator.”</li> </ul>	5 (13%)
<p><b>Staff were trained or educated</b> about evaluation. The evaluator’s goal was to help the staff become more independent or self-sufficient when</p>	5 (5%)	<i>n/a – No corresponding negative category</i>	

<p>working on the evaluation.</p> <p>Example:</p> <ul style="list-style-type: none"> <li>• “We were trained on the tool used.”</li> <li>• “They helped teach us more about the field. Their goal was to teach us to work on our own and not rely on her.”</li> <li>• “We had a random control trial started during 2011. This was a very positive experience during implementation as well as to elevate the level of understanding about evaluation and specifically outcomes of our work.”</li> </ul>	
<p><b>Participatory approach with a diversity of stakeholders.</b> The evaluator brought staff of different levels to the table.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “She encouraged us to include staff members from across the institution to discuss programs and how they affected everyone at every level. These discussions were comprehensive and meaningful.”</li> <li>• “The board and the entire staff were engaged at various points along the way. Everyone had a vested interest in the outcome.”</li> </ul>	<p>4 (4%)</p> <p><i>n/a – No corresponding negative category</i></p>



*Red Flags: Evaluation for Fundraising or Reputation*

Finally, I noted a few “red flags” in the comments: comments that nonprofits intended as being positive, but can actually be worrisome or negative aspects of an evaluation. Of the nonprofits that said their evaluation experience was positive, approximately one in ten (nine percent) mentioned a red flag. For example, one nonprofit was pleased that the evaluation process “confirmed what we suspected.” Table 7 describes “red flags” that appeared in the positive responses.

Table 7. Red flags

Positive Experiences	n (%)	Negative Experiences	n (%)
<p><b>Red flags:</b> The evaluation confirmed positive findings. Information was used for fundraising/marketing. Doing an evaluation set us apart from other nonprofits in the field.</p> <p>Examples:</p> <ul style="list-style-type: none"> <li>• “It gave more credibility to our findings. Additionally, working with an external evaluator set us apart from other nonprofits in our field.”</li> <li>• “Provided us the objective feedback that confirmed what we suspected.”</li> </ul>	8 (9%)	<i>n/a – No corresponding negative category</i>	

## **Chapter 5. Discussion**

There are 2.3 million nonprofit organizations in the United States (Roeger, Blackwood, & Pettijohn, 2011) and ninety percent of nonprofits are engaging in evaluation of some kind (Morariu et al., 2012). Multiple research projects have sought to better understand how and why nonprofits measure the results of their work (e.g., Alaimo, 2008; Carman, 2004; Carman, 2007; Carman, 2011; Carman & Fredericks, 2008; Carman & Fredericks, 2009; Carman et al., 2008; Carman & Millesen, 2004; Fine et al., 2000; Hoefler, 2000; Morariu et al., 2012; and Reed & Morariu, 2010, among others). Despite this prior research, little is known about nonprofits' reactions to the process of working with external evaluators. The purpose of this study is two-fold: 1) to understand how many nonprofits are working with internal and/or external evaluators and 2) to learn how they describe their experiences working with evaluators.

### **Purpose and Goals**

Three personal and professional goals guided this study. Namely, I sought to:

1. Uncover nonprofits' perspectives on their experiences working with evaluators;
2. Share these findings with evaluators to help them improve their consulting services to nonprofits; and
3. Increase funders' awareness and knowledge of their grantees' experiences with evaluation.

I completed the first goal: learning more about nonprofits' experiences with evaluation. The second and third goal are beyond the scope of this study, but can be achieved through future work and research.

### **Research Questions**

The current study was guided by two research questions: 1) How many nonprofits are working with internal and/or external evaluators? and 2) How do nonprofits describe their experiences with evaluation?

### **Methods**

These data come from a larger survey research project, the State of Evaluation 2012 study by Morariu et al. (2012). The full survey included thirty-eight questions about evaluation practices and capacity, of which twenty-five questions were closed-ended and five questions were open-ended. The full survey instrument is provided in the appendix. When conducting the State of Evaluation 2012 research, Morariu et al. (2012) contacted GuideStar ([www.GuideStar.org](http://www.GuideStar.org)) to purchase contact information for U.S.-based 501(c)3 Public Charities. The researchers were interested in organizations that had updated their IRS Form 990 in 2010 or more recently and had provided a contact name and email address. A total of 38,789 organizations met the criteria. In August 2012, Morariu et al. (2012) sent an initial email invitation to a random sample of 20,000 organizations, followed by three additional reminder emails to representatives who had not yet responded. The survey was open for twenty-five days. Throughout the survey administration period, logistical and technical issues prevented the researchers from reaching all 20,000 organizations. For example, several thousand surveys could not be

delivered because the receiving email server was busy and not receiving messages; because the receiving email inbox was full; because the receiving organization's firewall blocked the message as spam; or because the email address was incorrect due to a typo or other error. In several cases, Morariu et al. (2012) received automated messages that indicated that the organizations were no longer open. Out of 20,000 organizations, 13,070 organizations received the survey and 546 responses to the survey, for a response rate of four percent.

The current study is a deeper analysis of two survey items, Question 17 and Question 18/19. Question 17 asked, "If your organization worked with an external evaluator in 2011, what was your experience?" Respondents could select *horrible*, *poor*, *mixed*, *good*, *excellent*, *we did not work with an external evaluator*, or *I don't know*. Respondents who indicated that their experience was *horrible*, *poor*, or *mixed* were led to Question 18, an open-ended item that asked, "You indicated that you worked with an external evaluator in 2011 and it wasn't a great experience. Why do you think the experience wasn't positive?" Respondents who indicated that their experience was *good* or *excellent* were led to Question 18, an open-ended item that asked, "You indicated that you worked with an external evaluator in 2011 and it wasn't a great experience. Why do you think the experience was positive?" Nonprofits that responded *we did not work with an external evaluator* or *I don't know* were automatically skipped ahead to the next survey question.

## **Results**

*Research Question 1: How many nonprofits are working with internal and/or external evaluators?*

First, I examined how many nonprofits reported working with internal and/or external evaluators in the prior year (that is, 2011). Nine percent of nonprofits reported only working with their internal evaluator; eighteen percent worked only with an external evaluator; and seven percent worked with both internal and external evaluators.

*Research Question 2: How do nonprofits describe their experiences with evaluation?*

Thirty-three percent of survey respondents described their experience as *excellent*, 36 percent described their experience as *good*, twenty-seven percent described their experience as *mixed*, two percent described their experience as *poor*, and two percent described their experience as *horrible*. Overall, more than two-thirds of nonprofit respondents (sixty-nine percent) said their evaluation experience was either *excellent* or *good*.

The survey incorporated a skip pattern in which respondents who answered positively (excellent or good ratings) were led to an open-ended question that asked, “You indicated that you worked with an external evaluator in 2011 and it was a positive experience. Why do you think the experience was positive?” Respondents who answered negatively (mixed, poor, or horrible ratings) were led to an open-ended question that asked, “You indicated that you worked with an external evaluator in 2011 and it wasn’t a great experience. Why do you think the experience wasn’t positive?”

Ninety-one individuals provided open-ended responses about their positive experiences and forty individuals provided open-ended responses about their negative

responses. Positive explanations were slightly longer than negative explanations (a median of nineteen words compared to fifteen words).

Next, I coded the open-ended survey items. For example, one respondent who had a positive evaluation experience wrote, “The evaluator provided expertise in survey design and administration that we were unable to. The external evaluator name added validity to findings that an internal evaluation report could not.” The first sentence describes the evaluation consultant’s experience, knowledge, or expertise. The second sentence describes a different concept: the value of gaining a third-party perspective on their findings.

In total, I categorized the open-ended survey responses into twenty-two distinct positive codes and fifteen distinct negative codes. Positive codes are listed below, with corresponding negative quotes in parentheses as applicable.

1. The evaluator was experienced, skilled, competent, knowledgeable, professional, seasoned, prepared, or provided expertise. The evaluator brought more experience than was available in-house, which increased the nonprofit’s capacity. (Related negative code: The evaluator lacked expertise, competency, or professionalism.)
2. The evaluator possessed background knowledge of the organization, program, field, or content area. (Related negative code: The evaluator was unfamiliar or disconnected from the organization, the organization’s culture, or the programmatic area.)
3. Soft skills (e.g., good listener, excellent facilitator, organized, engaged, or committed). (No corresponding negative code.)

4. Value of a third party perspective: New, fresh, neutral, valuable, void of biases, another point of view; Not slighted by internal pressures. (Related negative code: Stressful to open up to an outsider.)
5. The evaluation process was useful, educational, or helpful and contributes to continued success. (No corresponding negative code.)
6. Met expectations; met needs, expectations and goals were clear, they did what we asked them to do, performed the required tasks; met deliverables. (Related negative code: Didn't meet expectations, or expectations weren't clear up front; perfunctory; unsatisfying experience.)
7. The evaluator guided the nonprofit to set up processes or think through new ideas. (No corresponding negative code.)
8. Appropriate depth and/or breadth. Holistic assessment; thorough. (Related negative code: Insufficient depth and/or breadth: Not as deep as they would have liked; or, too complicated.)
9. “Good” or tailored overall approach. Good approaches to evaluation. Tailored approach in which the evaluator adapts methods to fit nonprofit's unique needs. (Related negative code: Not tailored to nonprofit's unique needs. Inflexible. Evaluator focused on own needs.)
10. Resources: Effective and efficient use of resources, ample resources. (Related negative code: Limited funding. Challenges with pro bono or volunteer evaluator.)

11. Process was useful: The evaluation process was useful, educational, and helpful, and contributed to continued success. (No corresponding negative code.)
12. Findings were useful and used. The nonprofit gained new information or new insights about best practices. (Related negative code: Findings were disseminated or published without the nonprofit's permission.)
13. Excellent and practical recommendations. (No corresponding negative code.)
14. Methods and analysis: The evaluator developed measurable criteria, thoroughly analyzed data, and/or performed a meaningful analysis. (No corresponding negative code.)
15. Excellent and detailed reports. (Related negative code: Never received report.)
16. Adequate time and/or attention was given to the nonprofit or the evaluation process. The evaluator held regular meetings and may have met with the nonprofit whenever necessary. As a result, the evaluator was described as thoughtful, accessible, or responsive. (Related negative code: Insufficient time and/or attention was given. The evaluator may have been difficult to get a hold of. A few nonprofits felt the evaluator did not devote enough time to understand their organization.)
17. Good communication. (Related negative code: Poor communication.)
18. Deliverables were completed in a timely manner. (Related negative code: Deliverables were tardy; the process was slower than desired.)
19. Negative code: The evaluation was labor-intensive for the organization. (No corresponding positive code.)



20. Collaborative. The evaluation process was a joint effort between evaluators and nonprofits. There was close coordination as the evaluator and nonprofits worked together and exchanged ideas. (Related negative code: Not collaborative. The evaluators and nonprofits did not reach consensus on decisions throughout the evaluation planning or implementation. Results may not have been vetted or co-created.)
21. Staff were trained or educated about evaluation. The evaluator's goal was to help the staff become more independent or self-sufficient when working on the evaluation. (No corresponding negative code.)
22. Participatory approach with a diversity of stakeholders. The evaluator brought staff of different levels to the table. (No corresponding negative code.)
23. Red flags: The evaluation confirmed positive findings. Information was used for fundraising/marketing. Doing an evaluation set us apart from other nonprofits in the field. (No corresponding negative code.)

After completing this process, I grouped related codes together into broader categories. These concepts are complex and interrelated, and there are likely infinite different, and correct, ways to group these codes. One of my research goals was to make this information actionable for evaluators, that is, to gather information that could help evaluators improve their craft. I grouped the codes together based on 1) whether the respondent was describing the evaluation consultant or the evaluation process more broadly, and 2) by skillsets. The six parent codes included:

1. The Evaluator's Experience, Knowledge, and Soft Skills

2. Qualities of the Evaluation Process as a Whole
3. The Methods, Results, Reports, and Recommendations
4. Project Management: Timing and Communication
5. Collaboration (or Lack of Collaboration) with Nonprofit Staff
6. Red Flags: Evaluation for Fundraising or Reputation

## **Implications**

### *Percentage of Nonprofits Working with Evaluators*

This study found that nine percent of nonprofits are working exclusively with an internal evaluator and an additional eighteen percent are working exclusively with an external evaluator. In a few cases, nonprofits are working with both internal and external evaluators: seven percent of survey respondents indicated that they had worked with both internal and external evaluators in the year prior to the survey's administration (that is, in 2011). These numbers are much lower than in previous studies: Fine et al. (2000) found that twenty-nine percent of nonprofits were working with a combination of internal and external evaluators and Hoefler (2000) found that nearly three-fourths of nonprofits worked with both internal and external evaluators.

### *The Evaluator's Experience, Knowledge, and Soft Skills*

The first category was related to the evaluator's experience, knowledge, and soft skills. For example, twenty-five percent of nonprofits that had excellent or good evaluation experiences praised the external evaluator's skills, competence, and professional demeanor. An example is, "She provided expertise that we would have lacked otherwise." Similarly, twenty-eight percent of nonprofits that had mixed, poor, or

horrible evaluation experiences mentioned the external evaluator's lack of expertise. An example is, "Competency and professionalism issues."

*Qualities of the Evaluation Process as a Whole*

In addition to naming specific characteristics of the external evaluation consultant, nonprofit respondents also described characteristics of the evaluation more holistically. Specific codes in this category were related to the nonprofit's opinion of hearing a third party perspective of their organization, whether the evaluation met their expectations, and whether the evaluation provided the depth and breadth that was needed by their organization.

One common response, given by twenty-one percent of the nonprofits that had positive evaluation experiences, related to the perceived value of gaining an outside perspective on the organization's work. Respondents described this outside perspective as being new, fresh, neutral, valuable, void of biases, and not slighted by internal pressures. Interestingly, only three percent of nonprofits with negative evaluation experiences mentioned the third party perspective as a drawback of the process—primarily because it was "stressful" to work with an outside consultant. One person wrote "While we knew we needed a 'makeover' and major staff changes, it was stressful to open up to outsiders and admit the need."

Another area—mentioned by four percent of nonprofits with positive evaluation experiences but eighteen percent of nonprofits with negative evaluation experiences—was the match (or mismatch) between the overall evaluation approach and the nonprofit's needs. Those with positive experiences described the approach used as tailored, or simply

“good,” because the evaluator adapted methods to fit the nonprofit’s unique needs. For example, one person wrote that the evaluator was “willing to change and edit [the] evaluation tool so that it worked for our programs.” These comments speak to the value of spending time getting to know the nonprofit staff and the organization history and culture so that the evaluation can match their expectations and needs, rather than a predetermined theory or method that the evaluator may be accustomed to following. The “prescriptive nature” of evaluation is “difficult to achieve in the real world” explained evaluator Deb Rog at the Eastern Evaluation Research Society’s 2014 Eleanor Chelimsky Forum (Castillo, 2014). “Going by the book will always get you in trouble,” Rog warned the audience (Hamilton, 2014). Adapting evaluation designs and methods can be one of the most challenging skills for novice evaluators to learn, especially for those of us trained in the social sciences. Novice evaluators often approach their first few evaluation projects with “by the book” methodological practices (Piff, 2014), not realizing how much organizations can benefit from having evaluation designs and data collection instruments tailored to their information needs (Patton, 2011).

#### *The Methods, Results, Reports, and Recommendations*

This third category discusses the evaluation’s methods, results, reports, and recommendations. I grouped these topics together because they all relate to classic “textbook” skills that can be, for the most part, taught in training programs such as graduate coursework or conference workshops. This parent code includes five child codes about the utility of the evaluation process; the utility of the evaluation’s findings; recommendations; methods and analysis; and reports.

The first two codes—the utility of the evaluation process and utility of results—relate to *evaluation use* or *utilization*. Evaluation use is generally operationalized as the extent to which stakeholders use the information garnered throughout the evaluation to inform program structure, funding, staffing, policy change, and other components of decision making. Evaluators and stakeholders may enter into evaluation with a variety of perspectives about the evaluation’s broader purpose. Weiss (1973) observed that the rationale for conducting an evaluation is its usefulness in improving programs or informing policy decisions. According to the *Guiding Principles for Evaluators*, additional purposes and uses include “bettering products, personnel, programs, organizations, governments, consumers and the public interest; contributing to informed decision making and more enlightened change; precipitating needed change; empowering all stakeholders by collecting data from them and engaging them in the evaluation process; and experiencing the excitement of new insights” (“Preface: Assumptions Concerning Development of Principles,” n.d., para. 2). Essentially, there is interest in helping organizations benefit from the evaluation process itself and to use the findings to make improvements.

In 1986, Cousins and Leithwood broadly defined *evaluation results* “as any information associated with the outcome of the evaluation, for example, data, interpretations, recommendations; such information could be communicated at the completion of the evaluation or as the evaluation was proceeding” (p. 332). In 1997, Patton introduced the term *process use*, which he defined as the impacts that result from the learning occurring as a consequence of involvement in the evaluation process. Patton

wrote, “Process use refers to and is indicated by individual changes in thinking and behavior, and program or organizational changes in procedures and culture, that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process” (Patton, 1997, p. 90).

Both of these broad categories, *results use* and *process use*, appeared in the open-ended survey results. Of the nonprofits with a positive evaluation experience, ten percent noted the utility of the evaluation findings and another thirteen percent attributed their positive experience to the utility of the evaluation process. One survey respondent noted that “We have a report from which we have developed a number of committees to make improvements to the building and to the activities conducted. We have made some large steps in improving both areas of the evaluation” (utility of findings) and that “the process was educational and extremely helpful” (utility of process). *Results use*, in particular, is one of the most common uses and goals of nonprofit evaluation; nonprofits often engage in evaluation in order to improve or adjust their programming (Alaimo, 2008; Carman & Fredericks, 2008; Carman & Millesen, 2004; Fine et al., 2000; Hofer, 2000; Morariu et al., 2012). Survey respondents also noted another facet of evaluation use: evaluation misuse. Of the nonprofits that had negative evaluation experiences, one organization (3%) noted that the evaluation’s results were published and disseminated without their knowledge or permission.

Traditionally, the evaluation consulting engagement concludes with the delivery of a final report. In recent years, concerns over a “dusty shelf report” have prompted many evaluators to seek alternative reporting modes, such as verbal presentations,

webinars, slidedocs (standalone PowerPoint slides that are designed to be read as a story rather than to accompany a presentation), and dashboards, among others. Regardless of the delivery mode, the concept is the same: the evaluator or evaluation team condenses the quantitative and qualitative information collected throughout the evaluation into a cohesive report or presentation for the organization. Of the nonprofits that had positive evaluation experiences, four percent noted that they received “excellent” or “detailed” evaluation reports. “We received a detailed report that answered questions we set forth in the plan for evaluation,” described one of the survey respondents. Unfortunately, of the nonprofits that had negative experiences with evaluation, three percent said they never received a report. “We never got a full report and they kept insisting we needed to pay more money to get what we needed,” explained one of these nonprofits. Clearly it is important for evaluation consultants to discuss deliverables and project budgets throughout the evaluation to avoid a situation in which the nonprofit did not receive the report they were expecting.

Evaluation reports often contain a set of recommendations for improving the program or organization. These recommendations are often included within a final evaluation report or discussed verbally during meetings and presentations. Leviton and Hughes (1981) argued that providing practical recommendations increases the evaluation’s utility. Eight percent of the nonprofits with positive evaluation experiences attributed this experience to the excellent and practical recommendations made through the evaluation. For example, respondents indicated that the external evaluator “was able to provide feedback which enabled us to modify our work,” “made good suggestions for

future programming,” and that “their recommendations were practical” and the “input was invaluable.” In practice, very few of my own evaluation friends and colleagues include recommendations because they do not feel it is their role to advise the nonprofit on a specific course of action. Instead, they prefer to support the nonprofit throughout the learning process so that the organization’s leaders are equipped to make their own decisions about the findings. Perhaps providing recommendations within the final report is a practice worth revisiting.

#### *Project Management: Timing and Communication*

A key aspect of managing an evaluation is completing work within a timeframe that is useful for the organization. Only one percent of nonprofits with positive evaluation experiences mentioned that completing the project “in a timely manner” was a highlight of the experience. On the other hand, ten percent of nonprofits with negative evaluation experiences noted that the deliverables were tardy and/or the evaluation process in general was slower than desired. “Very slow to get results,” described one of the survey respondents. Completing work within a timely manner is typically a “given” of any consulting project, within the field of evaluation or otherwise; perhaps that is why so few nonprofits with positive experiences mentioned the timeframe at all. However, when evaluators miss deadlines or are generally slow to produce deliverables, nonprofits seem to take note of this flaw of the project management.

Evaluation consultants may miss deadlines for a number of reasons: due to understaffing issues such as staff illnesses or staff turnover, from being overcommitted and overscheduled from consulting with a number of different nonprofit clients



simultaneously, or due to unforeseen changes in the evaluation's design, such as delays in data collection. Even though many of these delays cannot be prevented, it is important for evaluation consultants to communicate changes in deadlines to nonprofits as soon as possible so that nonprofits are not surprised when reports are delivered later than originally planned.

*Collaboration (or Lack of Collaboration) with Nonprofit Staff*

Survey respondents with positive evaluation experiences described how the evaluation process was a collaborative, joint effort that involved close coordination and a good exchange of ideas. "The evaluation was a real collaboration between the researcher and the staff from start to finish," described one respondent. Another person said, "The evaluator and the staff worked together to create and implement the evaluation. It was a joint effort."

Perhaps these experiences went well because evaluators trained staff about evaluation: five percent of nonprofits with positive evaluation experiences mentioned that the evaluator(s) had provided training on evaluation. "They helped teach us more about the field. Their goal was to teach us to work on our own and not rely on her," explained a survey respondent. Other times, the evaluator(s) took a participatory approach and involved a diversity of stakeholders (four percent of the nonprofits with positive evaluation experiences). For example, a nonprofit explained, "[The evaluator] encouraged us to include staff members from across the institution to discuss programs and how they affected everyone at every level." Another person explained, "The board

and entire staff were engaged at various points along the way. Everyone had a vested in the outcome.”

*Red Flags: Evaluation for Fundraising or Reputation*

Finally, I noted a few “red flags” in the survey responses: comments that nonprofits intended as being positive, but can actually be worrisome or negative aspects of an evaluation. For example, one nonprofit was pleased that the evaluation process “confirmed what we suspected” (rather than being open to both positive and negative findings). Of the nonprofits that said their evaluation experience was positive, one in ten (ten percent) mentioned at least one red flag.

Several other organizations said their evaluation experience was positive because they were able to use the findings for fundraising and marketing activities (rather than for learning and improvement). These comments align with prior research results in which nonprofits report using evaluation to appeal to funders and potential funders, for outreach, for public relations, or to promote the program to potential participants (Alaimo, 2008; Carman & Fredericks, 2008; Carman & Millesen, 2004; Fine et al., 2000; Hoefler, 2000). For example, Hoefler (2000) found that, when describing their reasons for conducting an evaluation, forty-two percent of organizations wanted “to use the results for commercial reasons” and thirty-nine percent of organizations conducted an evaluation “because funders want one done” (p. 171). Carman (2011) found that “more than one-third of the nonprofit organizations in this study... conducted evaluations for monitoring or compliance purposes” (p. 365). She explained,

“By their own admission, these organizations are not capitalizing on the potential benefits of evaluation. Rather, they are just ‘jumping through the hoops’ and ‘reporting numbers’ to funders... The risk here is that normative and coercive pressures simply force organizations to adopt ritualistic behaviors that, in the end, devalue evaluations, and lend credence to the belief that evaluation drains organizations of valuable resources that could be better used to improve programs and services. The challenge with these organizations is to illustrate the value that evaluation information can have when used properly” (Carman, 2011, p. 365).

Another survey respondent attributed his or her positive evaluation experience to the fact that working with an external evaluator “set us apart from other nonprofits in our field.” Although none of these examples is egregious, they serve as a reminder that nonprofits choose to undergo evaluation for a number of reasons—for marketing, fundraising, and reputation-building—and not necessarily to learn about the outcomes of their work or to make data-driven improvements. Nonprofits and their funders invest too much money and staff time into evaluation to *not* benefit from the process and the potential for organizational learning. And, if nonprofits are viewing evaluation as a tool to boost their own reputation and visibility, are they missing an opportunity to use evaluation to improve their programming—and, therefore, to serve participants better and improve communities?

### *Topics Not Mentioned*

When asked to describe why their evaluation experience was positive or negative, the survey respondents noted a number of different reasons, such as the evaluator’s

knowledge, prior experience, and project management skills. Glesne (2011) advises qualitative researchers to “be cognizant of what was not said or demonstrated in some way” (p. 195). In the current study, survey respondents did *not* mention several common evaluation topics.

Logic models are diagrams that connect a nonprofit’s present services or activities to the outcomes or changes they are trying to achieve in the future. Logic models gained considerable popularity during the 1990s and 2000s through the United Way’s outcomes movement, and by the mid-2000s, approximately two in ten nonprofits were using a logic model (e.g., Carman, 2007; Carman & Fredericks, 2008; Carman & Millesen, 2004). By 2012, six in ten organizations used a logic model (Morariu et al., 2012). Despite their widespread adoption, none of the survey respondents mentioned logic models either positively or negatively.

Evaluators spend considerable time on the technical aspects of evaluation, such as refining data collection instruments and planning for and conducting data analysis. Evaluators also spend time preparing and finalizing deliverables such as handouts and reports, such as ensuring that documents are formatted nicely and free of jargon and typos. Again, none of these technical details of evaluation were mentioned directly by the survey respondents. Presumably, the evaluators that nonprofits had a positive experience with have already mastered these skills; that is, these aspects of the evaluation go “as expected” and do not stand out as noteworthy. However, these technical skills were not mentioned by any of the nonprofits with negative evaluation experiences either. An alternate explanation is that the technical side of evaluation--details of the instruments,

analysis, and report writing--are important to evaluators but are negligible or imperceptible to nonprofits, illustrating a difference in priorities between evaluators and nonprofit staff.

#### *Alignment with the Guiding Principles for Evaluators*

After coding the responses, I noticed several connections between the code hierarchy and the American Evaluation Association's (AEA) *Guiding Principles for Evaluators* (American Evaluation Association, 2004). The *Guiding Principles for Evaluators*—systematic inquiry, competence, integrity/honesty, respect for people, and responsibilities for general and public welfare—were developed by the American Evaluation Association “to guide the professional practice of evaluators, and to inform evaluation clients and the general public about the they can expect to be upheld by professional evaluators” (American Evaluation Association, 2004).

*Systematic inquiry* means that evaluators should conduct systematic, data-based inquiries about the program, policy, or initiative that is being evaluated. This principle also states that evaluators should “adhere to the highest technical standards appropriate to the methods they use,” “explore with the client the shortcomings and strengths” of evaluation questions and methods, and “communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret, and critique their work.” (American Evaluation Association, 2004, para. 12). A sample response that encapsulates this principle is, “They thoroughly analyzed our data and made corrections as needed; their reports were well written and thorough; their recommendations were practical.” A second example is, “We had a random control trial started during 2011. This

was a very positive experience during implementation as well as to elevate the level of understanding about evaluation and specifically outcomes of our work.”

The *competence* principle states that evaluators should provide competent performance to stakeholders. This principle is similar to the parent category I identified in the code hierarchy that relates to the evaluator’s experience, knowledge, and skills. A sample response from a nonprofit with a positive evaluation experience is, “Evaluator and her team were experienced and skilled in working with health organizations.” The *competence* principle also advises evaluators to “practice within the limits of their professional training and competence, and... decline to conduct evaluations that fall substantially outside those limits.” Practicing outside their own limits seemed to be one reason that evaluators received negative reviews. For example, one respondent described how “The evaluator seemed to be in over their head.” Another person felt that their evaluation experience was negative because the evaluator did “not [have] enough experience to accurately guide the final product.”

*Integrity/honesty* describes how evaluators should ensure the honesty and integrity of the evaluation process. One component of this standard discusses the need for open conversations about evaluation costs; the standard states that evaluators “should negotiate honestly with clients and relevant stakeholders concerning the costs... [and] it is primarily the evaluator's responsibility to initiate discussion and clarification of these matters, not the client's.” I noted these concerns in the *limited funding* code, which were expressed by one in five (twenty percent) of the nonprofits that had negative evaluation experiences. For example, a respondent wrote that “We never got a full report and they

kept insisting we needed to pay more money to get what we needed.” It is unclear whether the evaluator purposefully misled the nonprofit about costs (an honesty/integrity issue) or simply miscommunicated about costs (more of a negotiations or communications issue).

The *integrity/honesty* standard also states that, “if evaluators determine that certain procedures or activities are likely to produce misleading evaluative information or conclusions, they have the responsibility to communicate their concerns and the reasons for them. If discussions with the client do not resolve these concerns, the evaluator should decline to conduct the evaluation.” I noted this sentiment in my “red flag” code--warning signs in responses that could develop into situations that produce misleading evaluative information or conclusions. For example, one nonprofit attributed their positive evaluation experience to the fact that “working with an external evaluator set us apart from other nonprofits in our field.” While there is no evidence of an integrity/honesty issue, this mindset can lead to situations in which the nonprofit ignores undesirable findings or generally disengages from the evaluation process because their primary aim is to impress outside groups (rather than seeking to improve their services).

*Respect for people* specifies that evaluators should respect the security, dignity, and self-worth of stakeholders (e.g., respondents, program participants, clients, and other people with whom they interact). This standard further states that evaluators should “seek a comprehensive understanding” of contextual elements such as the study’s political and social climate, timing, and economic conditions. One survey respondent expressed this sentiment when he or she wrote that the evaluator “did not take the time to understand the

why we were at the level of functionality as an organization thereby not providing us good quality information,” which I coded as a mismatch or disconnect from the organizational culture (or social climate).

*Respect for people* also advises that evaluators should strive for social equity so that individuals participating in an evaluation can benefit in return (American Evaluation Association, 2004). A few organizations with negative evaluation experiences expressed the lack of social equity in their evaluation. For example, a survey respondent wrote that “it was very labor-intensive for us to get [the evaluator(s)] to fulfill their commitments.” This response illustrates the lack of social equity; rather than benefiting those who have given their time to the evaluation, the evaluation diminished a valuable resource--staff time.

The fifth standard, *responsibilities for general and public welfare*, states that evaluators should articulate and take into account the diversity of interests and values that may be related to the general and public welfare. For example, evaluators are instructed to include the full range of stakeholders and their viewpoints. I included these responses in a similar category called *participatory approach/encouraged diversity of stakeholders*. A sample response that encapsulates this standard comes from a nonprofit that reported having a positive evaluation experience: “[The evaluator] encouraged us to include staff members from across the institution to discuss programs and how they affected everyone at every level. These discussions were comprehensive and meaningful.”

*Is It Time for Evaluation Credentialing?*



Overall, nonprofits' perspectives of working with evaluation consultants are largely positive: more than two-thirds (sixty-nine percent) described their evaluation experience as *excellent* or *good*. When asked why they described their experiences as being positive, twenty-five percent of the nonprofits with positive experiences were complimentary about their evaluation consultant's prior experience, twelve percent ascribed their positive experience to their evaluation consultant's background knowledge, twelve percent discussed the evaluator's collaborative and inclusive approach to evaluation, and five percent noted their evaluation consultant's soft skills.

Although two-thirds of survey respondents had an *excellent* or *good* evaluation experience, the remaining one-third were not as positive: twenty-seven percent described their experience as *mixed*, two percent described their experience as *poor*, and two percent described their experience as *horrible*. When asked they described their experience as being *mixed*, *poor*, or *horrible*, survey respondents most commonly attributed their negative experiences to factors such as the evaluator's lack of experience (twenty-eight percent), insufficient background knowledge (twenty-five percent), ineffective use of resources (twenty percent), and inadequate time and attention devoted to the organization and project (twenty percent).

When implemented well, evaluation can be a powerful tool for both organizational accountability and learning. How can we ensure that more nonprofits have positive evaluation experiences and benefit from the resources—staff time and financial commitments—that they have devoted to the evaluation? Nearly half of nonprofits (forty-four percent) reported that knowing where or how to find an external evaluator was a

challenge (Morariu et al., 2012). If so many nonprofits are unsure where to find an evaluator in the first place, how will they know where to find a *good* evaluator?

Some evaluation skillsets can be learned in classroom settings during university coursework or formal conference workshops. Following university training, evaluators' knowledge and skillsets can be recognized with a degree (although only a few universities, such as Western Michigan University, specifically award a degree in evaluation). However, there is no recognition (i.e., no continuing education credits) for evaluators who complete full-day conference workshops. Other evaluation skillsets are learned on-the-job and can be recognized indirectly by listing relevant experiences on a resume or business proposal.

Some professions, such as medicine, law, and teaching, offer certifications to assure that individuals have undergone sufficient education and training before being allowed to practice their craft independently and without supervision. The Canadian Evaluation Society awards the Credentialed Evaluator (CE) certification, which supports “defining, recognizing, and promoting the practice of ethical, high-quality, and competent evaluation” (Canadian Evaluation Society, n.d.). Evaluators with a graduate degree and at least two years of evaluation experience may voluntarily submit applications to become credentialed. Evaluators must demonstrate that they are competent in the *Competencies for Canadian Evaluation Practice* and their applications are reviewed by the Credentialing Board (Canadian Evaluation Society, n.d.). The *Competencies for Canadian Evaluation Practice* are similar to the American Evaluation Association's

*Guiding Principles for Evaluators* in that both sets of competencies involve technical, interpersonal, and project management skills.

The American Evaluation Association (AEA), the primary professional association for U.S.-based evaluators, does not offer certifications to its members. Is it time for evaluation credentialing in the United States? Stevahn, King, Ghore, and Minnema (2005) argue that “program evaluation began as a field with a can-do attitude” (p. 43) but “in one area, however—the development of competencies for evaluators—the field of program evaluation has been decidedly less than can-do” (p. 44). Stevahn et al. (2005) describe the typical process for developing professional competencies: often, a group of practitioners generates a draft list of competencies, institute an expert review process to refine the competencies, and then update the competencies over time as theory, research, and practice advance. Evaluators have “followed a different course;” instead, evaluators have “debated... whether it was even possible to discover the common ground of competencies in an area of practice spanning such diversity across context and content” (Stevahn et al., 2005, p. 44). The resulting “chicken-egg situation” is such that, if there are no agreed-upon competencies for evaluators, then the field of evaluation can never become a unified profession and ensure that competencies are taught across training programs for novices or across continuing education programs for experienced evaluators (Stevahn et al., 2005, p. 44).

Stevahn et al. (2005) described several consequences of not having established competencies for evaluators. First, “incompetent evaluators, charlatans, and crooks may well pass as seasoned professionals” (Stevahn et al., 2005, p. 45). Second, hiring

processes are made more difficult without standard licensing procedures. Third, novices may have difficulty determining what they need to learn in order to become evaluators. Finally, individuals who design university training programs or professional development programs may base their programs on their own personal preferences rather than on unified, research-based competencies.

When describing the Canadian Evaluation Society's credential process, an evaluator wrote,

"...the CES process was introspective. I found it a valuable experience, it caused me to pause and think about what and why I do what I do as an evaluator, and I am proud of my CES designation. I am waiting for AEA to be brave enough to step up to the plate and do the same. Hopefully in my lifetime!" (Hwalek, 2013).

### **Limitations**

These data are drawn from a larger study (Morariu et al., 2012) in which researchers obtained 546 responses to an electronic survey. Both the larger study and the current analysis have several limitations.

#### *Limitations of the Larger Study*

There are several limitations of the larger State of Evaluation 2012 project. Survey responses may not be representative of the larger population of nonprofits due to low response rates, the survey respondent's role within the organization, and the organization's age and budget size.

The first limitation is the survey's relatively small sample size (and therefore, low response rate). These data are drawn from a larger study (Morariu et al., 2012) in which

researchers obtained 546 responses to an electronic survey. The sample is relatively small when compared to the larger population, and also had a relatively low response rate. The survey population includes 969,194 nonprofit organizations—the total number of U.S.-based 501(c)3 public charities that registered with the IRS in 2011, the year before the survey was conducted. A portion of these nonprofits—38,789 organizations—completed their IRS Form 990, indicating that they are currently open and in business. Then, researchers purchased a random sample of 20,000 email addresses from GuideStar.com; of these, 13,070 had valid email addresses. Overall, Morariu et al. (2012) calculated a response rate of four percent (546 survey responses divided by 13,070 possible responses).

Respondents worked in a variety of roles within their nonprofits, which may affect their view of the evaluator and the evaluation process. Most respondents (eighty-four) were executive staff, such as Executive Directors or Chief Operating Officers. Five percent were Board members, three percent were internal evaluators, three percent were program staff, and the remaining four percent held “other” roles. Each of these roles represents a different proximity to the evaluation processes that have taken place at their organizations, which might affect how they describe their organization’s experience working with an external evaluator. For example, an organization’s internal evaluator(s) is often closest to the evaluation process, serving as a liaison between the external evaluator and their coworkers. An internal evaluator would also have more knowledge and expertise about evaluation. On the other hand, Board member members are often farther removed from the evaluation process, receiving evaluation results but rarely

participating in day-to-day decisions or conversations as the evaluation process unfolds. Board members also have a unique perspective because they can observe multiple organization-wide processes simultaneously—for example, conversations about evaluation, strategic planning, budgeting, and other priorities—giving them a birds-eye-view of the evaluation. The current study did not examine whether the survey respondent's role affected their responses.

Organizations surveyed were generally older, and may have more evaluation experience than younger organizations: sixty percent of organizations were founded more than twenty years ago, twenty-one percent were eleven to twenty years old, twelve percent were six to ten years old, and the remaining seven percent were one to five years old. Older organizations are more likely to engage in evaluation than younger organizations (Morariu et al., 2012; Reed & Morariu 2010). The older organizations surveyed may have more evaluation experience and capacity than the full population of nonprofits nationwide.

Smaller nonprofits were underrepresented in the sample. Most nonprofits are small: Nationwide, nearly half of nonprofits (forty-six percent) have annual budgets below \$100,000 (Roeger et al., 2011). In this survey sample, however, only thirteen percent of respondents came from organizations with annual budgets below \$100,000. Smaller nonprofits are less likely to engage in evaluation than larger nonprofits (Morariu et al., 2012; Reed & Morariu, 2010). Additionally, an evaluation of a small nonprofit is very different from an evaluation of a large nonprofit. For example, a smaller nonprofit may operate a single program with a few dozen participants, so the evaluation may

consist of a survey or focus group of participants. A larger, multi-million dollar nonprofit may operate dozens of programs throughout a region, serving hundreds or thousands of people each year. That evaluation may involve more stakeholders, a variety of data collection methods, and a more rigorous design, such as a quasi-experimental or experimental design. Accordingly, smaller and larger nonprofits would likely report different experiences with their external evaluator and with the evaluation process as a whole. The current study did not examine whether smaller and larger nonprofits reported different experiences with evaluation.

Finally, these findings may be outdated. The State of Evaluation 2012 survey was administered during Summer 2012 and nonprofits were asked to report on their evaluation experiences during the prior year, 2011. At the time of this analysis, in Spring 2014, the findings are two to three years outdated. It is unclear whether any results from this study are still applicable or whether situations and contexts in the field have already changed. A future analysis could compare State of Evaluation 2010, 2012, and 2015 responses.

#### *Limitations of the Current Study*

There are several more limitations of the current study's design and analysis. For instance, while this analysis explores *how* nonprofits describe their experiences with evaluation, it does not answer *why*; that is, there are no connections between the coded responses and the nonprofit's demographic characteristics or other survey items. Future analyses could test whether the nonprofit's age, budget size, or other variables may influence their experiences with evaluation.

Additionally, the analysis is based on short open-ended survey responses. Participants' positive responses ranged from two to one hundred thirty-three words in length (median response length = nineteen words) and the negative responses ranged from two to sixty words in length (median response length = fifteen words). Open-ended survey items provide researchers with an opportunity to begin learning about the breadth of respondents' viewpoints, but do not provide the depth that can be gained from interviews, focus groups, observations, or other methods.

My coding was largely influenced by my experience as an evaluator. For the past few years I have provided evaluation guidance to nonprofits in both an internal and external evaluation role. A researcher with more experience, less experience, or no experience with nonprofit evaluation would likely interpret the responses differently. I discussed the general code hierarchy findings with the thesis advisor, but I was the sole coder.

A related limitation is that there were no member checks. Member checks are a strategy for improving the accuracy, credibility, and validity of a qualitative research or evaluation study. Member checks can take place during interviews or focus groups; the interviewer might summarize the participant's main points and ask the participant whether that summary is accurate or not. Member checks can also take place after data collection by sharing findings with participants and asking them to comment on whether their perspectives and viewpoints have been captured accurately in the findings. Member checks can also be used for quantitative studies. One application is participatory data analysis, in which evaluators share first drafts of findings with nonprofits and give



stakeholders an opportunity to provide input on the findings and recommendations (Pankaj, Welsh, & Ostenso, 2011). According to Pankaj et al. (2011), these member checks “draw on the wisdom of the people implementing the programs,” “make evaluation an opportunity for learning,” and reduce the chance that nonprofits view evaluation “as a punitive process that is ‘done to’ them” (p. 1). In future studies, researchers could share the findings of this study with nonprofit staff and ask them to comment upon the extent to which these findings resonate with their personal evaluation experiences.

Findings are based on nonprofits’ perceptions (for example, of the evaluator’s experience level), not on objective data about the evaluation process and the external evaluation consultant. Triangulation, a technique in which researchers collect data from two or more sources, is one strategy to increase validity of findings. When information that comes from several data sources or research participants—rather than from a single method or single type of research participant—converges into similar findings, researchers can be more confident that results are accurate. Future research could triangulate findings by collecting demographic data about the evaluation consultant(s) and the evaluation process.

A related limitation is that the survey respondents were, for the most part, not experts in evaluation: three percent of respondents identified themselves as evaluation staff, and the remaining 97 percent were executive staff, Board members, and program staff, among other roles. These non-evaluator survey respondents--the vast majority of respondents--may not be the best judge of the evaluation’s quality or of the evaluation

consultant's expertise. This analysis represents the respondents' perceptions of evaluation; the criteria for their judgments are not always evident. For example, one respondent who had a negative experience with evaluation attributed that experience to the evaluation consultant's "competency and professionalism issues." This brief, four-word explanation does not reveal what types of competency and professionalism issues the nonprofit encountered, or whether those complaints are reasonable concerns.

### **Suggestions for Future Research**

There are several opportunities for future research about nonprofits' experiences with evaluation that can both expand the scope of the current study and expand its limitations. First, the current study involved an analysis of open-ended survey responses which ranged in length from two to one hundred thirty-three words. Future studies could involve in-depth interviews, observations, document reviews, and/or other methods to gain deeper insights into these trends. Second, the current study is based on nonprofits' *perceptions* of the external evaluation consultant and of the evaluation process. Future studies could complement this subjective data with objective data. For example, researchers could ask nonprofits to provide demographic information about their evaluation consultant, such as the number of years that he or she has been working in evaluation. Future studies could also collect data on pairs of nonprofits and evaluators; that is, researchers could collect survey or interview data from both the nonprofit staff person and the external evaluation consultant. This research design would allow the researcher to determine how closely perspectives align—or don't—and what factors might lead to disagreements between nonprofits and evaluators.

## Appendix. Survey Instrument

1. Which category best describes your role within your nonprofit organization?
  - a. Executive staff
  - b. Program staff
  - c. Evaluation staff
  - d. Board member
  - e. Other, please specify: \_\_\_\_\_
  
2. What was your organization's zip code on December 31, 2011? \_\_\_\_\_
  
3. Which is the main geographic area in which your organization carries out its activities?
  - a. Internationally
  - b. United States
  - c. Regional level (more than one state)
  - d. State level
  - e. Local level
  
4. How old is your organization?
  - a. 1 – 5 years old
  - b. 6 – 10 years old
  - c. 11 – 20 years old
  - d. More than 20 years old
  
5. How many employees (including part-time and full-time employees, but not volunteers) did your organization have on December 31, 2011?
  - a. 1 – 3
  - b. 4 – 9
  - c. 10 – 19
  - d. 20 – 49
  - e. 50 – 99
  - f. 100 – 400

- g. 500 or more
6. What was your organization's annual budget for fiscal year 2011?
- a. Less than \$100,000
  - b. \$100,000 or more, but less than \$250,000
  - c. \$250,000 or more, but less than \$500,000
  - d. \$500,000 or more, but less than \$1 million
  - e. \$1 million or more, but less than \$2.5 million
  - f. \$2.5 million or more, but less than \$5 million
  - g. \$5 million or more, but less than \$10 million
  - h. \$10 million or more
7. In 2011, what was the biggest source of funding for your organization?
- a. Dues, fees, and other direct charges
  - b. Government grants or contracts
  - c. Charitable contributions from individual donors
  - d. Charitable contributions from foundations or philanthropies
  - e. Charitable contributions from corporations
  - f. Other sources (including interest from investments)
8. We define advocacy as "a wide range of activities conducted to influence decision makers at various levels." This includes not only traditional advocacy work like litigation, lobbying, and public education, but also capacity building, network formation, relationship building, communication, and leadership development. Approximately how much of your organization's work in 2011 (based on budget) was advocacy work?
- a. No advocacy work
  - b. Some advocacy work
  - c. About half advocacy work
  - d. Mostly advocacy work
  - e. All advocacy work
  - f. I don't know
9. How does your organization define evaluation? \_\_\_\_\_
10. In 2011, did your organization evaluate any part of its work (not including staff performance evaluations)?
- a. Yes

- b. No
- c. I don't know

11. In 2011, did your organization have at least one full-time employee dedicated to evaluation?

- a. Yes
- b. No
- c. I don't know

12. If you selected "no" or "I don't know" for question #11, who was primarily responsible for conducting evaluation work for your organization?

- a. Executive Director/President/CEO
- b. Evaluator(s) from outside the organization
- c. The organization's Board of Directors
- d. No one individual has the lead
- e. I don't know
- f. Other staff person, please specify title: \_\_\_\_\_

13. On average, how would you rate your organization's evaluation capacity?

- a. High internal capacity, for example, we regularly complete entire evaluations in-house
- b. Moderate internal evaluation capacity, for example, we do a good amount of evaluation work in-house, but sometimes we need help
- c. Low internal evaluation capacity, for example, we have a hard time doing things like tracking program data

14. In 2011, were the following groups a primary audience, secondary audience, or not an audience for your organization's evaluation work?

	Primary audience	Secondary audience	Not an audience
The organization's CEO/ED/management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other staff within the organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The organization's Board of Directors	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The organization's funder(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The organization's client(s)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other nonprofits in our field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Policy makers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15. In the past year, how has your organization used evaluation findings? (Check all that apply.)

- a. To make resource allocation decisions
- b. To make staffing decisions
- c. To plan/revise programs
- d. To plan/revise strategies
- e. To report to funder(s) on grants and/or contracts
- f. In communications and reports to stakeholders (other than clients)
- g. In proposals to funders
- h. To share best practices/lessons learned (e.g., presenting or publishing)
- i. In communications and reports to clients
- j. In updates/reports to the Board of Directors
- k. In support of advocacy/policy recommendations
- l. To implement scaling strategies

16. In the past year, how has your organization communicated your evaluation results?

- a. In reports to funder(s) on grants and/or contracts
- b. In proposals to funders
- c. In staff meetings
- d. In updates/reports to the Board of Directors
- e. In an annual report
- f. In an evaluation report
- g. On the organization's website
- h. Through the organization's social media
- i. In the wider media (newspapers, television, radio, etc.)
- j. In publications and presentations
- k. Other, please specify: \_\_\_\_\_

17. If your organization worked with an external evaluator in 2011, what was your experience?

- a. Horrible [Proceed to #18]
- b. Poor [Proceed to #18]
- c. Mixed [Proceed to #18]

- d. Good [Skip to #19]
- e. Excellent [Skip to #19]
- f. We did not work with an external evaluator [Skip to #20]
- g. I don't know [Skip to #20]

18. You indicated that you worked with an external evaluator in 2011 and it wasn't a great experience. Why do you think the experience wasn't positive? \_\_\_\_\_

19. You indicated that you worked with an external evaluator in 2011 and it was a positive experience. Why do you think the experience was positive? \_\_\_\_\_

20. Nonprofit organizations often have a document that communicates the relationship between the work completed by the organization and the changes that occur as a result of that work. Some common names for these types of documents are logic model, theory of change, program plan, log frame, etc. Does your organization have a logic model, theory of change, or other similar document for at least part of the organization's work?

- a. Yes
- b. No [Skip to #22]
- c. I don't know [Skip to #22]

21. In 2011, did your organization create or revise that logic model, theory of change, or other similar document?

- a. Yes
- b. No
- c. I don't know

22. Does your organization do the following before, during, or after beginning new programs/strategies?

	Before programming/ strategy implementation	During programming/ strategy implementation	After programming/ strategy implementation	We don't do this
Create a logic model or similar document	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Create an evaluation plan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Collect data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Analyze data

23. The following answer choices represent common evaluation designs—or how an evaluation project is set up. Organizations may use a combination of designs. It is unlikely that one organization would employ all of these. Please select the designs that your organization has used in the past year. (Select all that apply.)

- a. Strategic learning
- b. Developmental evaluation
- c. Outcomes evaluation
- d. Impact evaluation
- e. Formative evaluation
- f. Summative/retrospective evaluation
- g. Before and after measures
- h. Studies with control groups
- i. Quasi experimental research
- j. Randomized control trials
- k. Long term follow up with clients after they have stopped using your organization's services
- l. Economic evaluation, e.g., cost-benefit analysis, social return in investment (SROI)
- m. Other, please specify: \_\_\_\_\_

24. The following answer choices represent common practices related to evaluation. Organizations may use a combination of these practices. It is unlikely that one organization would employ all of these practices. Please select the practices that your organization has used in the past year. (Select all that apply.)

- a. Case studies
- b. Compiling statistics
- c. Completing grant reports
- d. Document review
- e. An evaluation workgroup or review board
- f. Feedback forms, questionnaires, or surveys
- g. Focus groups
- h. Internal tracking forms
- i. Interviews
- j. Structured observation
- k. Other, please specify: \_\_\_\_\_



25. Most evaluations typically seek to answer one or more of the following questions. Please rank the questions based on their importance to your organization during the past year. (With the response at the top being the most important.)

	1	2	3
How much? How many clients are served, how much service is provided, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How well? Were clients/participants satisfied, were the services provided high quality, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
What difference did it make? Were clients'/participants' lives changed, what changes did your work bring about, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

26. When evaluating your work, how important are the following?

	Very important	Somewhat important	Not very important	Not at all important
Learn whether original objectives were achieved	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learn about implementation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learn about outcomes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contribute to knowledge in the field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Strengthen organizational practices in the field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Strengthen public policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Strengthen future work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

27. Please rank in order of importance the following list of internal priorities that competed for resources in your organization last year (with the response at the top being the most important).

	Most important					Least important				
Communications	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Financial management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fundraising	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Governance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Human resources	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information technology	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Staff development	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Strategic planning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

28. This question asks how much your organization spent on evaluation in fiscal year 2011. In your estimate, please include all expenses spent in support of evaluation, expenses such as staff time, software for evaluation, evaluation consulting, contracts, etc. Approximately what percentage of your organization's budget was spent on evaluation in 2011?
- 0%
  - More than 0%, but less than 2%
  - 2% or more, but less than 5%
  - 5% or more, but less than 10%
  - 10% or more, but less than 15%
  - 15% or more, but less than 25%
  - 25% or more
  - I don't know
29. Which, if any, of your funding sources supported your organization's evaluation work in 2011? (Select all that apply.)
- 30.
- Dues, fees, and other direct charges [Skip to Question #31]
  - Government grants or contracts [Skip to Question #31]
  - Charitable contributions from individual donors [Skip to Question #31]
  - Charitable contributions from foundations or philanthropies
  - Charitable contributions from corporations
  - Other sources (including interest from investments) [Skip to Question #31]
  - None of our funders funded our evaluation work [Skip to Question #31]
31. Which, if any, of these corporate or philanthropic donor types funded your organization's evaluation work in 2011? (Select all that apply.)
- Nationally active funder, for example, Gates Foundation, Walton Family Foundation, Ford Foundation, Hewlett Foundation, etc.

- b. Locally active funder (that does not have a national presence)
- c. Community foundation
- d. Corporate funder
- e. United Way
- f. No foundations/philanthropies funded our evaluation work
- g. Family foundation
- h. Other, please specify: \_\_\_\_\_

32. In the nonprofit sector, the idea of scaling social impact refers to the increased ability of organizations to carry out their missions relative to the scope of the problem they address. This can mean serving more people and/or places, as well as deepening service to existing beneficiaries. An organization can scale by growing/spreading its entire organization, an individual program, a model, an idea or tool, a social movement, a policy initiative, a media or advocacy campaign, a marketplace solution, and so on. What have been your activities to date with regard to scaling the impact of your organization, program, initiative, etc.?
- a. Currently growing/scaling the impact of our organization, program, initiative, etc.
  - b. Interested in growing/scaling the impact of our organization, program, initiative, etc. but have not begun to implement a plan [Skip to Question #38]
  - c. We were previously engaged in growing/scaling the impact of our organization, program, initiative, etc. and are no longer doing so [Skip to Question #38]
  - d. We do not intend to grow/scale the impact of our organization, program, initiative, etc. [Skip to Question #38]

33. Please describe your organization's mission. \_\_\_\_\_

34. Please describe your organization's scaling initiative. \_\_\_\_\_

35. What is your strategy to grow/scale your impact? (Please select your highest priority.)
- a. Replicating our program or model in other geographic locations
  - b. Expanding our target populations by providing our current services to new people
  - c. Increasing the quality, quantity, or frequency of our services to current people being served
  - d. Promoting our model for adoption by other organizations
  - e. Lobbying or conducting advocacy campaigns to influence public policy
  - f. Establishing or expanding a social movement that influences public opinion, behavior, and/or collective action in support of our goals and mission
  - g. Other, please specify: \_\_\_\_\_

36. As part of that strategy, what are you focused on growing/scaling? (Please select your highest priority.)
- Our organization as a whole
  - A program
  - An idea or innovation
  - A technology, product, or service
  - A social movement
  - A public policy initiative
  - Other, please specify: \_\_\_\_\_

37. Does your organization do the following before, during, or after beginning a new scaling initiative?

	Before beginning a scaling initiative	During scaling initiative	After scaling initiative	We don't do this
Create and/or refine a logic model	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Create and/or refine an evaluation plan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Plan for data collection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Collect data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Analyze data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

38. What is the role of evaluation in your scaling efforts? \_\_\_\_\_

39. Please indicate the degree to which each of the following challenges inhibited your organization's ability to conduct evaluation in 2011:

	Not a challenge	Minor challenge	Significant challenge	n/a
Insufficient financial resources	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Limited staff time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Limited staff expertise in evaluation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insufficient support from organization leadership	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Funders asking you to report on the wrong data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not having staff who believe in the importance of evaluation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Knowing where or how to find an external evaluator	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

40. How do you feel about these statements?

	Strongly agree	Agree	Disagree	Strongly disagree	n/a
“I regularly discuss evaluation findings with funders.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“When I discuss evaluation findings with funders, I find the experience useful.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“Funders tend to ignore evaluation results in their decision making.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“You don’t need evaluation to know that your approach is working.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“There is too much pressure on nonprofit organizations to measure their results.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“Funders are accepting of failure as an opportunity for learning.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“Data collection interferes with our relationship with clients.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
“Most data that is collected in my organization is not used.”	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## References

- Alaimo, S. P. (2008). Nonprofits and evaluation: Managing expectations from the leader's perspective. In J. G. Carman & K. A. Fredericks (Eds.), *Nonprofits and evaluation. New Directions for Evaluation*, 119, 73-92. doi:10.1002/ev.269
- American Evaluation Association. (2004). *American Evaluation Association Guiding Principles for Evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). *Qualitative analysis on stage: Making the research process more public*. *Educational Researcher*, 31, 28-38.
- Behrens, T. R. & Kelly, T. (2008). *Paying the piper: Foundation evaluation capacity calls the tune*. *New Directions for Evaluation*, 119, 37-50. doi:10.1002/ev.267
- Blackwood, A. S., Roeger, K. L., & Pettijohn, S. L. (2012). *The nonprofit sector in brief: Public charities, volunteering, and giving, 2012*. Retrieved from Urban Institute website: <http://www.urban.org/UploadedPDF/412674-The-Nonprofit-Sector-in-Brief.pdf>
- Brock, A., Buteau, E., & Herring, A. (2012). *Room for improvement: Foundations' support of nonprofit performance assessment*. Retrieved from the Center for Effective Philanthropy website: <http://www.effectivephilanthropy.org/assets/pdfs/Room%20for%20Improvement.pdf>
- Carman, J. G. (2004). *Explaining program evaluation practice*. A paper presented at the Southeastern Conference for Public Administration held October 3 at the Hilton, University Place, Charlotte, NC.
- Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation*, 28, 60-75. doi:10.1177/1098214006296245
- Carman, J. G. (2011). Understanding evaluation in nonprofit organizations. *Public Performance & Management Review*, 34, 350-377. doi:10.2753/PMR1530-9576340302

- Carman, J. G. (2011). What you don't know can hurt your community: Lessons from a local United Way. *Nonprofit Management & Leadership*, 21, 433-448. doi:10.1002/nml.20035
- Carman, J. G. & Fredericks, K. A. (2008). Nonprofits and evaluation: Empirical evidence from the field. In J. G. Carman & K. A. Fredericks (Eds.), *Nonprofits and evaluation. New Directions for Evaluation*, 119, 51-71. doi:10.1002/ev.268
- Carman, J.G., & Fredericks, K.A. (2009). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation*, 31, 84-104. doi:10.1177/1098214009352361
- Carman, J. G., Fredericks, K. A., & Introcaso, D. (2008). Government and accountability: Paving the way for nonprofits and evaluation. In J. G. Carman & K. A. Fredericks (Eds.), *Nonprofits and evaluation. New Directions for Evaluation*, 119, 5-12. doi:10.1002/ev.265
- Carman, J. G. & Millesen, J. L. (2004). *Evaluation theory and practice: A report from the field*. A paper presented at the 6th International conference of the International Society for Third Sector Research held July 12 at Ryerson and York University, Toronto, Canada.
- Castillo, I. [isaac\_outcomes]. (2014, April 28). Deb Rog makes great point - prescriptive nature of #eval difficult to achieve in real world. Need flexibility and rigor balance. #eers14 [Tweet]. Retrieved from [https://twitter.com/isaac\\_outcomes/status/460778220371972096](https://twitter.com/isaac_outcomes/status/460778220371972096)
- Flaherty, M. P. & Stephens, J. (2013, November 24). Md. Woman accused of stealing millions from D.C. nonprofit group. *The Washington Post*. Retrieved from [http://www.washingtonpost.com/investigations/prosecutors-say-maryland-woman-embezzled-51-million-from-dc-nonprofit-group/2013/11/24/97e1ecf0-53bb-11e3-a7f0-b790929232e1\\_story\\_1.html](http://www.washingtonpost.com/investigations/prosecutors-say-maryland-woman-embezzled-51-million-from-dc-nonprofit-group/2013/11/24/97e1ecf0-53bb-11e3-a7f0-b790929232e1_story_1.html)
- Fine, A. H., Thayer, C. E., and Coghlan, A. (2000). Program evaluation practice in the nonprofit sector. *Nonprofit management and leadership*, 10(3), 331-339. doi:10.1002/nml.10309
- Fournier, D. M. (2005). Evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Glesne, C. (2011). *Becoming qualitative researchers: An introduction*. Boston, MA: Pearson Education, Inc.
- Grantmakers for Effective Organizations and TCC Group (2011). *Patterns and trends in*

*grantmaking: A national field study of grantmaker practices—2011*. Retrieved from the Grantmakers for Effective Organizations website: [http://docs.geofunders.org/?filename=2011\\_field\\_study\\_full\\_data.pdf](http://docs.geofunders.org/?filename=2011_field_study_full_data.pdf)

Greenlights for Nonprofit Success (2013). *Measuring nonprofit impact: The state of evaluation in central Texas*. Retrieved from Greenlights for Nonprofit Success website: [http://www.greenlights.org/resources/resource-library/doc\\_download/170-measuring-nonprofit-impact-the-state-of-evaluation-in-central-texas](http://www.greenlights.org/resources/resource-library/doc_download/170-measuring-nonprofit-impact-the-state-of-evaluation-in-central-texas)

Hamilton, J. [limeygrl]. (2014, April 28). Deb Rog - "going by the book will always get you in trouble" advice to new evaluators at #EERS14 #eval [Tweet]. Retrieved from <https://twitter.com/limeygrl/status/460786165743108096>

Hendricks, M. (2000). Outcomes measurement in the nonprofit sector: Recent developments, incentives, and challenges. A paper prepared for the American Society for Public Administration, Center for Accountability Performance, Symposium on Leadership of Results Oriented Management in Government, February 11-12, 2000, George Washington University.

Hendricks, M., Plantz, M. C., & Pritchard, K. L. (2008). Measuring outcomes of United Way-funded programs: Expectations and reality. In J. G. Carman & K. A. Fredericks (Eds.), *Nonprofits and evaluation. New Directions for Evaluation*, 119, 13-35. doi:10.1002/ev.266

Hoefler, R. (2000). Accountability in action? Program evaluation in nonprofit human service agencies. *Nonprofit Management and Leadership*, 11(2), 167-177. doi:10.1002/nml.11203

Hoole, E. & Patterson, T. E. (2008). Voices from the field: Evaluation as part of a learning culture. In J. G. Carman & K. A. Fredericks (Eds.), *Nonprofits and evaluation. New Directions for Evaluation*, 119, 93-113. doi:10.1002/ev.270  
Internal Revenue Service, 2007, p. 4

Hwalek, M. (2013, November 12). Certification, example from market research association [Electronic mailing list message]. Retrieved from <https://listserv.ua.edu/archives/evaltalk.html>

Kern, T. (2013). *The utilization of program evaluation data by nonprofit organizations* (Unpublished master's thesis). DePaul University, Chicago, IL.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage Publications.



- Morariu, J., Athanasiades, K., & Emery, A. (2012). *State of evaluation 2012: Evaluation practice and capacity in the nonprofit sector*. Washington, DC: Innovation Network.
- Morley, E., Vinson, E., & Hatry, H. P. (2001). *Outcome measurement in nonprofit organizations: Current practices and recommendations*. Waldorf, MD: Independent Sector.
- Murphy, D.M., & Mitchell, R. (2007). *Building evaluation capacity in North Carolina's nonprofit sector: A survey report*. Raleigh, NC: Institute for Nonprofits.
- Nonprofit Technology Network (2012). *The state of nonprofit data*. Retrieved from the Nonprofit Technology Network website: <http://www.nten.org/research/2012-state-of-data>
- Pankaj, V., Welsh, M., & Ostenso, L. (2011). *Participatory data analysis: Expanding stakeholder involvement in evaluation*. Washington, DC: Innovation Network, Inc.
- Patton, M. Q. (2011, November). *Utilization-focused evaluation*. Pre-conference workshop presented at the annual meeting of the American Evaluation Association, Anaheim, CA.
- Patton, M. Q. (2013, October). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Pre-conference workshop presented at the annual meeting of the American Evaluation Association, Washington, DC.
- Piff, J. [4socialimpact]. (2014, April 28). Biggest mistake Deb Rog sees among new evaluators: Going by the book! Need to ask how you can make a difference during evaluation. #EERS14 [Tweet]. Retrieved from <https://twitter.com/4socialimpact/status/460786389445910530>
- Preface: Assumptions Concerning Development of Principles. (n.d.). In the American Evaluation Association Guiding Principles for Evaluators. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Reed, E. & Morariu, J. (2010). *State of evaluation 2010: Evaluation practice and capacity in the nonprofit sector*. Washington, DC: Innovation Network.
- Roeger, K. L., Blackwood, A., & Pettijohn, S. L. (2011). *The nonprofit sector in brief: Public charities, giving, and volunteering, 2011*. Retrieved from Urban Institute website: <http://www.urban.org/publications/412434.html>
- Salamon, L. M., Sokolowski, S. L., & Geller, S. L. (2012). *Holding the fort: Nonprofit*

*employment during a time of turmoil*. Retrieved from the Johns Hopkins University website: [http://ccss.jhu.edu/wp-content/uploads/downloads/2012/01/NED\\_National\\_2012.pdf](http://ccss.jhu.edu/wp-content/uploads/downloads/2012/01/NED_National_2012.pdf)

Saldana, J. (2013). *The coding manual for qualitative researchers*. Retrieved from [http://www.sagepub.com/upm-data/24614\\_01\\_Saldana\\_Ch\\_01.pdf](http://www.sagepub.com/upm-data/24614_01_Saldana_Ch_01.pdf)

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation*, 26, 43-59. doi: 10.1177/1098214004273180

United Way of America (1996). *Measuring program outcomes: A Practical Approach*. Alexandria, VA: United Way of America.

United Way of America (2000). *Agency experiences with outcome measurement*. Retrieved from Innovation Network website: <http://www.innonet.org/resources/node/697>

United Way of America (2003). *Outcome measurement in national health & human service and accrediting organizations*. Retrieved from United Way website: <http://studio.unitedway.org/CILresources08/files/program%20outcomes/Outcome%20Measurement%20in%20Health%20and%20human%20services.pdf>

W. K. Kellogg Foundation (2004). *W. K Kellogg Foundation evaluation handbook*. Battle Creek, MI: W. K. Kellogg Foundation.

## **Biography**

Ann K. Emery is an evaluator for foundations and nonprofits in Washington, DC. At Innovation Network, she evaluates initiatives for the Annie E. Casey Foundation, Kansas Health Foundation, JPB Foundation, Center for Community Change, National Parks Conservation Association, and the Holocaust Memorial Museum. Prior to joining Innovation Network, she was an internal evaluator at the Latin American Youth Center and an educational researcher at ICF International. She serves on the Board of the Washington Evaluators, the Advisory Board of the Potent Presentations Initiative, and as the Co-Chair of the Data Visualization and Reporting group within the American Evaluation Association. She received her Bachelor of Arts in Psychology and Spanish from the University of Virginia in 2008.