The Role of Evaluations in Community Foundations

Brad R. Watts
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Policy Design, Analysis, and Evaluation Commons, Social Statistics Commons, and the Social Work Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/478

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE ROLE OF EVALUATIONS IN COMMUNITY FOUNDATIONS

by

Brad R. Watts

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Interdisciplinary Ph.D. in Evaluation
Advisor: David Hartmann, Ph.D.

Western Michigan University
Kalamazoo, Michigan
December 2011
The role of evaluations in community foundations

Brad R. Watts, Ph.D.
Western Michigan University, 2011

Each year, U.S. community foundations are responsible for billions of dollars in philanthropy, yet the outcomes associated with these efforts often remain unknown. Previous research supports the importance of evaluating philanthropic activities and shows that community foundations express a strong interest in evaluation; however, the limited available evidence suggests that evaluation practice is still not widespread.

This study reports the findings from a national survey of community foundations on evaluation practice. The findings indicate that a substantial share of community foundations do not formally evaluate the outcomes of their philanthropic work. Additionally, although previous research has suggested that community foundation evaluation practice is constrained by limited organizational resources, an analysis of asset and community size found that these traits are not associated with most measures of evaluation practice. Although community foundations with larger endowments do show a somewhat greater tendency to make funding decisions based on evaluations and to invest in evaluation-related training for staff, they did not show a consistently higher tendency to actually practice evaluation. Overall, the analysis suggests that barriers to evaluation related to organizational size or assets may have been overstated in the past.

Finally, this study also looks at alternate reasons why community foundations may or may not choose to practice evaluation. Research conducted with a subset of
community foundations offers additional support for the notion that factors besides assets or size play a role in the adoption of evaluation practices. Among community foundations that are active evaluators, internal characteristics including staff leadership and board influence were cited as deciding factors for overcoming barriers to evaluation.
ACKNOWLEDGMENTS

I would like to extend a warm thank you to everyone that has been a part of my graduate studies up through this point. As an interdisciplinary student, much appreciation must be given to all the students and faculty that welcomed me into their classes and shared with me a bit of their discipline and perspective. I would also like to recognize the staff of the Evaluation Center and my fellow students in the IDPE, who ultimately formed the core of my graduate school experience. Finally, I am always grateful to the Upjohn Institute for providing me with employment for so many years, motivating my desire to pursue graduate work, and serving as a second graduate school of sorts.

It should also be acknowledged that much of the burden of a dissertation is borne by the committee. A special thanks to Dr. David Hartmann and Dr. Kevin Hollenbeck, for your support and guidance, and to Dr. Michael Scriven for starting me down this path in the first place. Finally, much love and appreciation to Jody for tolerating me throughout the years and for providing another set of eyes on this final work.

Brad R. Watts
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii
LIST OF TABLES ................................................................................................................ vi

CHAPTER

I. INTRODUCTION ........................................................................................................... 1
   History and Role of Community Foundations......................................................... 2
   Problem Statement .................................................................................................. 5

II. EXISTING LITERATURE ON CF EVALUATION ..................................................... 8
   So Why Evaluate? The Value Argument .............................................................. 18

III. METHODOLOGY .................................................................................................... 21
   Survey of Community Foundations ....................................................................... 22
      The Survey Instrument ....................................................................................... 23
      Population and Sample ....................................................................................... 23
      Survey Response ................................................................................................. 24
      Survey Sample Size ............................................................................................ 27
      Possible Sources of Bias ..................................................................................... 29
   Testing for Bias Using Known Characteristics .................................................... 30
   Other Possible Sources of Bias ............................................................................. 35
   Survey Data Collection and Analysis .................................................................... 38
      In Depth Data Collection from Select Community Foundations ...................... 38
   Selection of Participants and Bias ......................................................................... 39
   Qualitative Methodology ....................................................................................... 39
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis Approach and General Hypothesis</td>
<td>40</td>
</tr>
<tr>
<td>IV. SURVEY RESULTS</td>
<td>42</td>
</tr>
<tr>
<td>General Survey Results</td>
<td>42</td>
</tr>
<tr>
<td>Analysis for Differences by CF Characteristics</td>
<td>59</td>
</tr>
<tr>
<td>Survey Response Differences of CFs with Large Endowments</td>
<td>67</td>
</tr>
<tr>
<td>Survey Response Differences of CFs with Large Service Area Populations</td>
<td>76</td>
</tr>
<tr>
<td>Findings from the Analysis by Endowment and Service Area Population</td>
<td>83</td>
</tr>
<tr>
<td>IV. QUALITATIVE ANALYSIS</td>
<td>86</td>
</tr>
<tr>
<td>Methodology</td>
<td>86</td>
</tr>
<tr>
<td>Triangulation Support for Previous Findings</td>
<td>88</td>
</tr>
<tr>
<td>Follow Up Participant Perspectives on Evaluation Practice and Use</td>
<td>92</td>
</tr>
<tr>
<td>Internal Factors Drive Evaluation Practice</td>
<td>92</td>
</tr>
<tr>
<td>Breadth of Evaluation Practice Exceeds Depth</td>
<td>94</td>
</tr>
<tr>
<td>Barriers to Evaluation</td>
<td>95</td>
</tr>
<tr>
<td>Resources for CF Evaluation</td>
<td>100</td>
</tr>
<tr>
<td>Summary of Follow Up Interview Findings</td>
<td>103</td>
</tr>
<tr>
<td>VI. CONCLUSIONS</td>
<td>105</td>
</tr>
<tr>
<td>Connection of Literature and Research Findings</td>
<td>105</td>
</tr>
<tr>
<td>Theoretical Hypotheses and Study Findings</td>
<td>108</td>
</tr>
<tr>
<td>Contribution to Literature</td>
<td>110</td>
</tr>
<tr>
<td>Lessons for Community Foundations</td>
<td>112</td>
</tr>
<tr>
<td>Suggestions for Future Research</td>
<td>114</td>
</tr>
<tr>
<td>Lessons Learned</td>
<td>116</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>119</td>
</tr>
</tbody>
</table>
APPENDICES

A. Survey of Community Foundations and Evaluation Usage................................. 125
B. Survey Follow-Up......................................................................................... 131
C. HSIRB Approval Letter................................................................................ 133
D. Census Bureau Regions and Divisions with State FIPS Codes....................... 135
LIST OF TABLES

1. Distribution of CFs by Census Division................................................. 32
2. CF Endowment Descriptive Statistics.................................................. 34
3. In-Place Systematic Procedures to Conduct Evaluations....................... 44
4. Level of Encouragement or Requirement for Grantee Evaluation Practice.................................................. 45
5. Ratings of Evaluation Usage by CFs for Selected Purposes...................... 47
6. Recommendation or Promotion of Evaluation Guidelines by CFs.............. 49
7. Percent of CF Respondents by Reported Rate of Use of Inclusion of Related Evaluation Practices................................................. 52
8. Percent of CFs Rating the Importance of Select Organizational Activities.................................................. 53
9. Questions on CF Staff Training in Evaluation Practice.......................... 56
10. Questions on CF Evaluation Promotion and Usage.................................. 57
11. CF Endowment and Service Area Size.................................................. 59
12. Test of Independence by Endowment and Population Category.................. 62
13. Question 2: Percent of Responses by Endowment Category...................... 68
14. Question 5 Part 1: Percent of Responses by Endowment Size Category...... 70
15. Question 5 Part 2: Percent of Responses by Endowment Size Category...... 71
16. Question 6 Part 2: Percent of Responses by Endowment Size Category...... 72
17. Question 7: Percent of Responses by Endowment Size Category.............. 73
18. Question 8: Percent of Responses by Endowment Size Category.............. 74
19. Question 11: Percent of Responses by Endowment Size Category.............. 75
List of Tables – Continued

20. Question 2: Percent of Responses by Service Area Size Category ................. 78
21. Question 3 Part 8: Percent of Responses by Service Area Size ..................... 79
22. Question 7: Percent of Responses by Service Area Size Category .................. 80
23. Question 8: Percent of Responses by Service Area Size Category ................. 81
24. Question 10: Percent of Responses by Service Area Size Category ............... 82
25. Question 11: Percent of Responses by Service Area Size Category ............... 83
CHAPTER I

INTRODUCTION

This study looks at the role that evaluation plays in community foundations (CFs) in the United States today. Although a full body of literature exists on both what CFs do and the ways CFs can add value to philanthropy, limited research has been developed to determine the degree to which evaluation is actually used by CFs. The studies of CF utilization of evaluation that have been done so far indicate that formal evaluation is seldom used, yet at the same time proclaim that interest in evaluation is growing (see for example Pauly, 2005). This is not surprising, given that formal evaluation procedures are relatively new to the world of philanthropy, having originally come forth primarily in the field of education during the 1960s, 1970s, and even the 1980s through the work of academics such as Campbell, Cronbach, Scriven, and Weiss—researchers whose methodological and philosophical works form the basis of the modern field of evaluation today (see Shadish, Cook, & Leviton, 1991). Still, it suggests that the acceptance of evaluation by CFs may still be in a state of flux, where both the current status and future standing of evaluation to CFs remains relatively unknown.

The rest of Chapter I provides an overview of the history of CFs and clarifies the problem addressed by the study. Chapter II details what existing literature is able to say about the use of evaluation by CFs, the challenges to expanding use of evaluation, and the arguments developed for and against increased CF evaluation practice. Next, Chapter III explains the research methodology employed to examine the degree to which CFs use
evaluation. The findings from the research are presented in Chapters IV and V: Chapter IV presents the findings from a quantitative analysis of the survey data and Chapter V covers the outcome of a qualitative interview process. Finally, Chapter VI summarizes the overall findings of the study, the implications for CFs, and suggestions for future research and evaluation activities.

**History and Role of Community Foundations**

Community foundations first entered the world of philanthropy with the creation of the Cleveland Foundation in 1914. The concept arose from an idea put forth by Fredrick H. Goff, a local banker and lawyer, to combine the assets of local philanthropists into a single trust that could be used to serve residents of the community (Cleveland Foundation, 2010). Funds from donors would be placed under the control of a single organization, which would be responsible for monitoring the endowment and distributing funds to individuals or charitable organizations in the community. Previously, many small philanthropic endowments had been managed by the banks, which lacked expertise in identifying how best to distribute funds from numerous small family and community funds; by combining these funds into a single separate corporation directed by a committee of community experts, Goff hoped to improve the efficiency of philanthropy in the Cleveland area (Marshall County Community Foundation, 2010).

The new community foundation form of philanthropic trust differed from common existing forms of philanthropy, such as corporate foundations, family foundations, religious-based charities, and non-profit organizations by focusing on endowment creation and management for a specific geographic region. In many ways, community foundations filled a niche between traditional church-based charities, which
addressed local concerns but were limited in outreach by religious conditions and small assets, and large private foundations, which often possessed large assets but were typically nationally-focused and constrained to addressing the social issues of most concern to their funders. Early CFs concentrated their efforts on a local area, though the initiatives were often still guided by the few bankers, donors, and community leaders that were responsible for the foundation’s founding. However, as independent organizations, the CFs soon became more responsive to community needs and implemented pioneering efforts to assess community needs, including developing some of the earliest community surveys, which were conducted in the early 1920s (Minter, 2008).

The concept took off quickly and by the 1920s CFs had been established in most major U.S. cities (Marshall County Community Foundation, 2010). However, multiple challenges caused growth to stall during the 1930s and 1940s. The Great Depression and World War II both limited the number of donors, while a 1935 change in tax law led more businesses to establish their own corporate foundations instead of endowing the local CF. Although CFs began to slowly grow in asset size and number again during the post-war era, it was not until the late 1960s that rapid growth resumed and the modern era of CFs began.

A major driver of CF growth was the Tax Reform Act of 1969, which clarified the types of charitable foundations under the tax code and the deductibility of donations (Internal Revenue Service, 1984). The new tax reform made contributions to CFs more attractive to donors by both increasing regulations on corporate foundations, while at the same time providing a higher maximum deduction threshold for donations to CFs and allowing CF donors to claim the full value of non-cash donations such as real estate. In
the 1970s, growth at CFs was also helped by an increase in marketing the CF concept directly to communities, instead of just through banks, and through the invention of new options for donors, such as donor advised funds (Carson, 2002). Later still, the conditions of the 1980s contributed to CF growth through both increasing prosperity and increasing need. Asset wealth increased during the decade, while at the same time cuts in federal government spending gave rise to new demands for privately-funded social programs to address the reductions in public programs (Marshall County Community Foundation, 2010).

Today, the importance of CFs as a means of charitable activity has increased dramatically. In recent years, the wealth and charitable grant-making of CFs has expanded rapidly, which has vastly increased their importance as funders of nonprofit organizations and charitable activities in the U.S. Today there are more than 700 CFs in the U.S. controlling assets of $56 billion. Even after accounting for inflation, total annual giving by CFs has increased by more than 600 percent since 1990. In 2008, CFs in the U.S. distributed over $4.6 billion in grants to individuals and non-profit organizations—a level of giving that now represents one out of every 10 charitable grant dollars distributed nationwide. (Foundation Center, May 2009)

As an organizational format, CFs have evolved to become major players in the charitable world over a relatively short period of time. However, increases in endowment assets or the number of organizations may not be enough to sufficiently increase the capacity of CFs to serve their communities. With an increase in assets and an increase in need also comes an increase in the importance of both identifying the most important need and making the most efficient and effective funding decisions possible. As the next
section discusses, there remains a gap in knowledge regarding what steps CFs are taking to monitor and direct their own activities.

**Problem Statement**

Given the relative collective size of CFs and the general increase in demand for accountability in the nonprofit sector, one might expect evaluation to be a common part of CF activities. Yet, surprisingly, little evidence exists to suggest that CFs are consistently engaged in the practice of evaluating their grantees. A search conducted using Google Scholar revealed a relative dearth of research on CFs in general and none of the first 400 search engine hits returned a scholarly reference to CF evaluation practices.\(^1\) Additional searches conducted using more traditional academic sources also produced similar results, with the majority of search hits referring to CFs only because of their role as a source of funding or support.\(^2\) While these searches are neither definitive nor comprehensive of all research on CFs and evaluation, the disappointing size and relevance of the results does highlight the comparative scarcity of research compared to other fields.

Additionally, research that does address foundations and evaluation usage often has not focused specifically on CFs, but taken a broader approach that does not take into account the unique nature of CFs. For example, McNelis and Bickel (1996) addresses

---

\(^1\) A Google Scholar search conducted on January 14, 2010 using the term “community foundations” returned 8,120 results. Sorted in order of relevance, the first 400 were individually examined, out of which only 16 covered subject directly related to community foundations while the remaining 384 search returns consisted primarily of research studies or articles that referred to community foundation funding or program support. A search conducted narrowing the domain to “‘community foundations’ + evaluation” narrowed the results to 4,120 but was similar in that the top hits were not articles specific to CFs, but usually covering broader topics.

\(^2\) A search conducted January 18, 2010 using ProQuest (via wmich edu) returned only 41 scholarly journal references for the term “community foundation” of which only 10 articles consisted of some form of research about CFs.
evaluation use across all type of foundations, not just CFs, while Carman’s (2007) article addresses evaluation use with a focus on community-based grantee organizations. Both examples do contribute some to the knowledge of CFs and evaluation usage, however, neither directly addresses the topic of concern to this study.

Of course, a lack of publicly available evaluation reports or scholarly reports on CF evaluation practices does not indicate that evaluation is not occurring in these organizations. To the contrary, evaluation does occur in CFs; however, the practice certainly varies and the overall quality of these evaluations and the nature of its usage remains relatively unknown. Most CFs must practice some form of evaluation when reviewing funding applications and boards make evaluative decisions when they take actions such as approving or rejecting programs. Assessment of the major outputs of CFs, however, may or may not be occurring in a formal sense.

Another issue of concern is evaluation usage. Even if CFs conduct formal evaluations of grantee programs or other funding activities, the value of the evaluation will be limited if the results are not used in decision-making or available to the relevant stakeholders. Conversations with the staff of the Kalamazoo Community Foundation, for example, suggest that evaluation efforts do occur at many CFs but it is often focused on internal use in management or funding decisions (Erway & Hoekstra-Terrentine, June 30, 2009). Use of evaluation for board or staff decisions is a clear and important step for CFs to take, however, if results are kept internal it prevents the scale of evaluation from being known and may limit potential benefits to donors, grantees, and community members.

Although the typical level of evaluation practice and usage remains unknown, many have become concerned that CFs are not evaluative enough. When Seita (1993)
examined the subject, he too found that while applications for funding were widely subject to systematic evaluation by CFs, the outcomes of activities funded by the CFs were not. Pauly’s (2005) newer work suggests that little has changed and makes a straightforward claim that “most foundation activities are not evaluated” (p. 9). However, the real problem may be that the level and usage of evaluation by CFs is largely unknown.

Without systematic data on evaluation in CFs, the overall breadth and depth of the practice remains a mystery, with little evidence to support the claims of those concerned by a lack of evaluation or those claiming strong levels of practice and use. To move beyond anecdotal evidence and theoretical discussion of how evaluation plays a role in CF philanthropy, research must be undertaken to address the major questions about the current role of evaluation in CFs. As such, this study will address the following research questions: 1) to what degree do CFs currently practice evaluation; 2) how do CFs that practice evaluation use the results; and, 3) are differences in evaluation practice or use associated with organizational characteristics? Developing an empirically-based answer to these questions can benefit the field by both providing an argument for CFs to engage in rigorous evaluation activities, as well as to provide a better understanding of the degree to which CFs currently excel at, or are lacking in, evaluative practices.
CHAPTER II

EXISTING LITERATURE ON CF EVALUATION

The lack of systematic evidence about the evaluative practices of CFs as an overall group should not be taken as a sign that evaluation is absent from the work of all CFs. At a minimum, “all foundations conduct some form of due diligence before making a grant” (Hofland 2007, p.57), although the criteria used may vary widely between organizations. In some form or another, all CFs must implement some sort of application process for individuals and nonprofit organizations that are seeking grants. Even if a CF is distributing funds that are marked for a very specific activity or a limited number of possible recipients, there must generally be in place some sort of information collected to determine that basic criteria are met: e.g. does the organization serve the arts or will the individual scholarship recipient be attending nursing school. In addition to assessing whether grant applicants meet basic criteria to receive funding, it will often be the case that CFs must also assess the relative quality of grant applicants, since the level of demand for grants is usually higher than the level of funding that CFs make available during a granting cycle.

An issue of greater concern, however, regards both the quality of evaluation practice and the degree to which evaluation is used to benefit those who have a stake in the activities of CFs. The aforementioned basic evaluative review of grantee criteria represents a minimum that will naturally occur during the process of administering grant funds in any manner that is short of a random, lottery-type distribution. There are
numerous approaches to grant application review, ranging from simple forms to review panels, which can vary greatly in both complexity and quality.

The review of grant applications is not the only aspect of CF activity where evaluation can or should play a role. Although when viewed externally it is perhaps the most obvious point in the process of distributing funds where CFs engage in assessment, as this research will show, it is not even the evaluation activity that is most widely recognized by CFs. Evaluation is perhaps most frequently thought of as a measurement of outcomes that occurs after funds have been awarded and a program has been implemented or an activity has occurred. Still, because some CFs do not conduct or demand outcome evaluations, as Ostrower (2007) and others suggest is the case, the initial application for funds may in some instances be the one and only chance some CFs have to evaluate grantees.

This process of evaluating funding proposals typically entails reviewing competing grant applications and making or denying awards based on an assessment of the applicant based on how they fit with the CF’s priorities, the apparent feasibility of the proposal, and the general strength of the applicant in terms of their leadership, financial structure, and overall organizational capacity (Hofland, 2007). In some cases, the process of grant application evaluation may be merely a formality of paperwork review, while in other cases it may attain the ideal of being a rigorous and systematic process, but at the very least it represents a minimum extent to which all CFs can be expected to report engaging in evaluation.

The second form of evaluation that CFs might be expected to engage in is the summative or outcome evaluation of grantee activities. According to the Joint
Committee’s (1994) evaluation standards, this type of evaluation is “designed to present conclusions about the merit or worth of an object and recommendations about whether it should be retained, altered, or eliminated” (p. 209). In the context of CFs, this is the process by which they evaluate how well the programs and organizations they fund are performing in the community. This type of evaluation can also be applied to the CF itself through the evaluation of the performance of the organization. As the review of literature presented in this section shows, outcome evaluation is not universally performed by CFs.

There are several reasons for not performing outcome evaluations, one being that CFs find it difficult to define success and measure outcomes in the community (Ostrower, 2007). In addition to the straightforward challenges of identifying tools or sources of data for measuring outcomes, there is also the fact that the needs of the community change over time, which can alter what achievements can be considered successful. While family or corporate foundations are free to focus on only a few pet issues selected by the board or members of the donor family, CFs should ideally assess any outcomes based on the impact on the specific needs of the community. This means identifying the needs, either through informal channels or by using a formal needs assessment process, which can be both expensive and time consuming.

Additionally, community needs may be both deep and general in nature, which makes it difficult to identify outcomes during a typical grant-making cycle. For example, poverty or poor health are both common community-level problems that can be difficult to impact in a measurable way over a short period of time. Identification of the full effects of a program that attempts to address a long-term issue can be hampered by the
inability to commit to tracking measures over time or difficulty in ensuring that future changes have a causal relationship with program outputs that occurred in the past.

Since doing good for a broad community often entails delivering a social impact that can only be achieved through delivering information and transforming existing systems, creating recognizable outcomes can be a time-consuming process (Kramer, 2005). The activities of the organizations and programs that CFs select to fund should be evaluated on a more frequent basis; however, this limits the evaluation to measuring outputs that have a theoretical tie to positive outcomes in the future, since the outcomes themselves are potentially too abstract or are expected to occur too far into the future to be easily measured, especially in a quantifiable manner.

Still, interest in evaluation, or at least talk of interest in evaluation, of the organizations and activities supported by CFs has greatly expanded over the last two decades (Pauly, 2005). This should be a good sign for CFs, since on the surface it would seem to suggest that the benefits of systematic evaluation are being recognized. However, at the same time, previous literature has already suggested that most foundations, including CFs, are not conducting nor using evaluation at the expected level (for example see Carson, 2000; Pauly, 2005; Seita, 1993; Ostrower, 2004, 2006; McNeilis & Bickel, 1996). The disconnect between the documented level of interest in evaluation and the documented level of evaluation practice and use by CFs leads one to the question of why this gap exists and whether there is any evidence that it is changing today.

One reason for the disconnect may be the short-term costs that evaluation may impose on CF staff. Since CFs are primarily grant-making organizations with limited
ability to directly impact social issues on their own, the outcomes generated by grantees are typically identified as a key measure of their own organizational effectiveness (Ostrower, 2004). As such, bad grantee evaluations may be taken by some as reflecting on the decisions made by CF staff. Therefore, CF interest in evaluation may be tempered by a view of evaluation as a risk, since the effectiveness of the staff and organization itself may end up being judged on the outcomes of grantee organizations that are not entirely within their control. Although there may be a gain to the CF from learning if a grantee program is good or not, that evaluation benefit will be weighed against the perceived risk of possibly coming up with evidence that money was spent on an unsuccessful program, which could potentially affect community support and donations. Additionally, staff may fear that evaluation results could have internal consequences as well, such as if the board believes that negative evaluation results for programs funded by the CF reflect on staff decision making or grantee screening processes.

Another reason that CFs may not evaluate is the expense of performing a rigorous evaluation of grantee projects. These expenses are essentially the easily identifiable direct costs of conducting evaluation, such as staff salaries or fees paid to consultants who conduct evaluations, as opposed to the more complex indirect costs that are cited throughout the literature as being barriers to evaluation. Although Alkin and Ruskus (1984) suggest that it should be possible to calculate all evaluation costs, they also admit that a “simple dollar-savings criterion is obviously inadequate in those instances where cost economies cannot be demonstrated” (p. 15). For CFs, determining the benefits of the evaluation of a complex social program may be seen as too challenging or outside the
mission, while the direct costs will be easily quantifiable and may seem high, particularly for a well designed and rigorous evaluation.

Overcoming the cost hurdle may also seem daunting if the monetary value is low, as can be the case for CFs that issue small grants or scholarship awards. The cost of an evaluation will certainly outweigh the benefit if viewed in the context of a program that awards one or two scholarships on an annual basis; however, when considered in the light of a awarding in perpetuity a small grant that is not effective, the benefit of conducting an evaluation will be far more likely to outweigh the cost. Still, for CFs it may be difficult to justify expending limited funds toward evaluation for small grant amounts, particularly if the funds are directed at one-time community projects. In other words, a program that awards two annual $1,000 scholarships may justify an evaluation, since the program intervention will ultimately expend a large sum of funds over a long period of time, while a one-time grant of $2,000 to support a school program or summer art fair may not warrant a full evaluation, since the cost of the evaluation would be likely to exceed the grant amount and the program would be unlikely to repeat regardless of the outcome.

One middle approach is to foster the development of evaluative capacity in organizations that receive funds, as opposed to the CF directly conducting evaluations or requiring contracts with outside evaluation consultants. Existing research on the evaluation conducted by CFs suggests that some have chosen to focus on providing tools that can increase the evaluative capacity of grantee organizations, not on conducting or requiring actual evaluations of funded activities (Graddy & Morgan, 2006). Capacity building can entail activities such as offering training to grantees on evaluation techniques, providing access to databases or other data resources, or through additional
financing for outside evaluation expertise. By helping grantee nonprofit organizations learn how to track outcomes and conduct basic evaluations, CFs can provide a long-term benefit; however, the downside is apparent in the form of self-evaluator bias and limited transfer of evaluation results back to the CF or other community stakeholders. In short, grantee organizations face very real incentives to produce glowing evaluation reports and to limit access to their internal data if they perceive negative results as threatening future funding opportunities.

The expression of interest in evaluation and efforts to bolster the capacity of the organizations doing the footwork of addressing community issues is certainly a step forward, however, it is not the same as actively engaging in evaluation. What evidence there is on the subject of CF evaluation practice suggests that the majority still are not practicing formal evaluation, despite growing interest and recognition of its importance. Across all types of foundations, the Urban Institute found only 44 percent conducted formal evaluations of funded activities in 2003 (cited in Pauly, 2005). Ostrower’s (2006) study found that the proportion of CFs that characterizes evaluation as very important was slightly lower than for other types of foundations: 40 percent versus 45 percent. Of larger concern, however, was the finding that fewer than half of those CFs that identified evaluation as “very important” actually engaged in formal evaluation practice: only 44 percent (Ostrower 2006). This suggests that external factors may be one element, along with obvious internal constraints such as funding limitations, that has been hampering evaluative practice in CFs despite the interest of organizational leaders.

In addition to the issues identified by academic researchers, practitioners in the field have identified downsides of evaluation as well. Some CF staff cite their reluctance
to engage in evaluation of grantees as being tied to the issue of utility, which is to say the findings of evaluation studies do not adequately serve their decision-making needs or provide the data necessary to improve the organization's practices. For example, Carson (2000) discussed the case of the U.S. Governmental Accountability Office's expression of frustration with research on educational programs, such as Head Start, that fail to draw clear conclusions on efficacy and instead recommend only that further research be conducted. If the results of an evaluation do not set a clear course of action for the CF leadership with regard to the project being examined, then the organization may feel that scarce resources have been wasted by engaging in an evaluative process. Unfortunately, past examples of poor evaluations may have tainted the opinion of some CF leaders.

Of course, the argument that evaluation is not always useful, timely, or precise in its findings is nothing new and certainly not unique to CFs. However, if the issue was simply a matter of improving the quality of evaluations being conducted, then employing existing tools such as Scriven's (2007) Key Evaluation Checklist, Stufflebeam's (2007) CIPP model, and the long-held utilization-focused approach advocated by Patton (2002) could help mitigate the claim of a lack of usefulness as a barrier to evaluation practice by CFs. The more persuasive claims, however, suggest that CFs avoid evaluation because the downside of evaluation is viewed as being potentially larger and more troublesome than the benefit to the organization.

According to Hall (2003), the real challenge to evaluation in a foundation setting is that the organizations are essentially ownerless, with interests split unevenly between donors, management, and the community. Stakeholders typically have limited power to influence the direction of the CF; only the CF's board, and to some extent the
government through regulation and tax policy, has any real authority over how the organization is operated and whether or not any sort of evaluation of funded activities is conducted. To CFs, engaging in evaluation may seem like all risk and no reward, since “even [evaluation’s] proponents concede it is expensive and time-consuming” (Hall, 2003, p. 3), and negative evaluation results can result in the ire of donors and the public. Simply put, the stakeholders impacted by CF programs have little power to make sure that rigorous evaluation is conducted, while CF leadership may fear that inconclusive or negative findings will make the organization look bad, while positive findings will not necessarily increase support for the organization nor satisfy all stakeholders.

The problem of staff concern with evaluation practice can be seen even in organizations that do conduct and use evaluation. Since some may be concerned that evaluation results showing that a program funded by the local CF was a failure could create negative publicity for the organization and generate pressure on the staff not to make mistakes in the future, there becomes an incentive to limit access to the findings. Not surprisingly, research indicates that among those CFs that do conduct or require formal evaluation procedures, very few make the results available to donors or the public (Pauly, 2005). Instead, most CFs restrict evaluation findings to internal uses, such as strategic planning and making decisions on future grant awards.

While CF leadership and staff may view the restriction of evaluation as an effective way to shield the organization from excessive criticism while still gaining evidence about grantee effectiveness, it instead may both limit the usefulness of the evaluation and generate anxiety among the staff about the findings. For example, restricting evaluation findings to internal use effectively limits the upside benefits that
could arise from a negative evaluation report, such as the opportunity for program improvement or increasing knowledge about approaches to solving social ills. At the same time, restricting evaluation to internal use still maintains a downside for the staff, who may face repercussions for making the decision to fund an activity that does not perform well.

CF staff may also be skeptical of program evaluation since there is little evidence to suggest that proving the success of a program through outcome evaluation actually results in increasing or securing financial support for that program or activity (Carson, 2000). In short, good evaluations may not always bring donor or community support. In fact, the ability of a CF to attract donations that will support its grantmaking activity is more closely associated with factors not related to the outcomes of its grantees, such as the age of the CF in the community, the level of trust and civic engagement in the community, and the socioeconomic characteristics of the community’s residents (Graddy & Wang, 2009). Therefore, a CF with a stellar track record of selecting grantees that operate effective programs to address pressing community needs may still not be able to build financial support for additional granting activities if the community lacks the necessary level of income or engagement to grow the donor base. Conversely, programs that address popular topics at CFs that are located in wealthy or growing regions may find donors lining up to support initiatives despite little or no evidence of success.

The final major form of evaluation that CFs could be expected to engage in is the evaluation its own operations. Like any other organization or activity—such as the grantees and their activities that CFs sometimes do evaluate—the workings of the CF itself can also benefit from evaluation. Beyond learning about the effectiveness of its
grantees, CFs that self-evaluate can gain a formal understanding of how the organization learns from the experiences of grantees, assess how well the organization sets priorities, know more about what has been accomplished, and even potentially share knowledge to advance practice at other CFs (Slater, Constantine, & Braverman, 2004). Self or internal evaluation by CFs is simply an extension of the aforementioned major types of expected evaluation—evaluation of funding requests and evaluation of grantees and grantee activities—that could be used to both improve the organization and promote its effort to community and donor stakeholders.

However, self evaluation, like other forms of evaluation, is still not widely practiced by CFs. One reason for this is that the dual role of CFs, which serve both donors and community, creates an environment where stakeholder needs may differ or even conflict. With different stakeholders potentially disagreeing on what constitutes effectiveness, measures of organization-wide effectiveness are not easy to identify and, unlike other types of organizations or programs, the relevance of measures may vary between individual CFs because of the large variation in the communities served. Additionally, all the downsides associated with the evaluation of grantees also apply to overall CF organization evaluation: high costs, uncertain outcomes, and resistance by staff who may fear poor results. Not surprisingly, these issues have resulted in a situation where CFs report feeling torn between interests and shy away from actually conducting evaluations, even as they acknowledge its potential importance. (Ostrower, 2006)

So Why Evaluate? The Value Argument

Given the potential downsides to evaluation and the lingering resistance to the practice of rigorous evaluation, at least amongst a large portion of CFs, it is important to
consider what the benefits to increased use of evaluation could be to these organizations. In a general sense, there is a philosophical argument for evaluating nearly all activities, including the activities of CFs and the grantees they fund. Besides fulfilling legal obligations or accountability requirements, evaluation results should provide benefits to the organization or activity being evaluated by identifying strengths, as well as pointing out areas that need improvement. Additionally, good evaluation should improve the efficiency of society in general by identifying those activities that do not produce results or that cannot produce results in a cost-efficient manner compared to alternative uses.

In the case of CFs, there are more specific pro-evaluation arguments as well. Pauly (2005) lists the basic uses of evaluation for all types of foundations as follows:

- Evaluations identify important and useful lessons reflecting the varied context and diverse goals of foundations.
- Evaluations provide practical information on ways for foundations, their grantees, and others to obtain improved results.
- Evaluations counterbalance the all-too-human tendencies for foundation staff (like other leaders) to indulge in delusions of success.
- Evaluations build reliable and useful evidence about the performance and the effects of the innovations supported by foundations. (p. 9)

The case made by Pauly focuses on utility and improvement for foundations; however, for CFs there are additional constituencies that also make a case for conducting evaluations. This forms the basis for the argument that evaluations should be considered a source of additional value that CFs can provide to grantees and communities.

One criticism of CFs is that they have become mere charitable bankers that are focused more on maintaining their own endowments and catering to new donors than serving the community. This can be seen in the move of CFs away from unrestricted
funds and toward donor-advised funds that can only be spent on specific types of issues or programs, as well as in a shying away from controversial social problems (Carson, 2002). Additionally, the endowment focus is reflected in the fact that endowments have in the past continued to grow, even through times of economic recession, while the amount of funds dispersed has stayed very near the legal minimum of five percent (Irvin, 2007). Critics have now begun to more frequently argue that CFs should more aggressively spend down their endowments to address current problems, focus more on community needs and less on administering narrow donor-advised funds, and find additional ways to benefit the community beyond simply holding and distributing funds.

One potential way CFs can avoid these criticisms and increase the value of activities to the community is through evaluation. Unlike individual donors or small charitable organizations, large CFs have the resources and expertise to evaluate grantees and direct resources towards those that can be the most effective, which can both improve the efficiency of the return on each charitable dollar and send a signal to other donors outside the CF system regarding a program’s effectiveness. Additionally, CFs could provide value by offering non-cash technical assistance to grantees and conducting long-term research and evaluation studies that could provide generalizable results on the effectiveness of a program approach. (Porter & Kramer, 1999)
CHAPTER III

METHODOLOGY

The research methodology for this project seeks to answer the following primary questions about the use of evaluation by CFs: 1) to what degree do CFs currently practice, 2) how do CFs use evaluation, and 3) does evidence exist that evaluation is associated with specific traits of a CF. Put another way, this study looks at how CFs evaluate and whether evaluation usage rates can be changed or if it is a fixed aspect of most CF organizations.

A first step in addressing these questions was to review recent literature on the topic of CFs and evaluation, which was presented in the previous chapter. On the question of evaluation practice by CFs, previous studies have already suggested that the majority of CFs do not regularly engage in rigorous, high quality, summative evaluations (see for example Ostrower, 2004, 2005, 2006; Pauley, 2005; Seita, 1993). However, these studies do not reflect the most current conditions and have focused on either broader samples (such as charitable foundations of all types) or narrower samples (only CFs within a small geographic region) instead of undertaking a categorical examination of all CFs. This study addresses these issues by utilizing a nationwide sample of organizations specifically designated as CFs and by revisiting the topic of practice and usage through multiple quantitative and qualitative measures.

The literature has also built support for the importance of evaluation, particularly Porter and Kramer’s (1999) argument that includes evaluation as a key component of
providing increased value to communities and Scriven’s assertion that evaluation should always provide a return greater than its cost (cited in Alkin & Ruskus, 1984). Those who have documented the non-use of evaluation have also gone on to suggest that current levels of evaluation practice are suboptimal (for example, Pauly 2005, Ostrower 2006). The key suggestion that drives this research is therefore the concept that CFs should gain substantial benefit from evaluation. Still, solid evidence on the quality of evaluation in those CFs that do regularly engage in evaluation and whether positive impacts from the practice are felt in the CF organization or the community remains scarce. Since there is a logical construct for why CFs should evaluate, it leaves the questions of whether evaluation actually remains as underused as others have suggested and if so, why?

To address these lingering questions, further research into the CF evaluation practice is required. Unfortunately, a review of the literature, as well as major foundation websites, indicated that no regular consistent source of data on CF evaluation is currently available. Therefore, this study relies on primary data collection to obtain new information on CFs and evaluation. To accomplish this, a two-step data collection process was developed: first a survey of CFs, followed more in-depth interviews conducted with a small subset of survey participants. The next two sections of this chapter detail the design and implementation of the survey and the follow-up interview process. Finally, the last section of this chapter discusses the analysis process, as well as general hypotheses for the research findings.

Survey of Community Foundations

Information on CFs and their use of evaluation were collected through a mail survey. Although more expensive than an email survey, a mail survey was selected
because of the availability of contact information through the 2010 Foundation Directory Online and because a higher response rate was expected. A mail survey also offers the benefit of allowing the respondent to look up data (Czaja & Blair, 2005). A telephone survey was ruled out because of both cost and time concerns. Ultimately, a mail survey was determined to offer the best opportunity to collect an unbiased, nationally representative sample of CFs.

The Survey Instrument

The survey instrument was developed solely by the researcher and designed to answer the first research question regarding the degree and quality of CF evaluation practice. The instrument contained 15 questions and was designed to fit on a single 11 X 17 sheet of paper, which was folded to fit into a #10 envelope along with a postage paid return envelope addressed for delivery to the Evaluation Center at Western Michigan University. An introduction at the top of the survey explains the purpose of the research, provides a disclaimer regarding the limitation of risks to the participant, and offers the researcher’s contact information in case the potential respondent has any questions or concerns about the survey. The survey instrument was reviewed by WMU’s Human Subjects Institutional Review Board (HSIRB) and determined to be exempt from oversight. A copy of the full survey instrument and a letter of approval from the HSIRB are provided in the appendix.

Population and Sample

An initial query of the Foundation Directory Online professional edition using the free public access showed listings for 1,369 organizations potentially classified as CFs by the Foundation Center (2010). However, upon obtaining access to the full database
system through the Kalamazoo Public Library, it became clear that the true universe of CFs was smaller; a query using the full professional version of the database allowed the elimination of organizations that did not report assets, did not report making grants, or reported being part of another type of organization. These non-active and non-standard organizations were excluded from the population so as to limit the universe of possible respondents only to CFs operating in a traditional environment and with granting experience that could be subject to evaluation. The final result was a list of 888 organizations that for the purpose of this study have been classified as CFs. This number is slightly larger, but within the range of other published estimates of the number of CFs in the country.

Because the population size was originally expected to be larger than 888, the decision was made to survey the entire population instead of conducting a random sample of a smaller subset of CFs. This allowed the advantage of providing all CFs with an equal opportunity to respond to the survey, as well as theoretically increasing the total number of responses that were likely to be received, which is one way of increasing survey reliability.

**Survey Response**

The survey was mailed to all 888 CFs on the mailing list during the last week of September 2010. Replies were requested to be placed in the mail by October 15, 2010; however, the acceptance of returned surveys was kept open for more than a month, until November 24, 2010, in order to allow for late return mailings and possible post office delays.
In total, 139 completed surveys were returned and only one survey was accompanied with an indication that the organization did not exist and could not be surveyed. This resulted in a response rate of 15.7 percent. The response rate was lower than hoped, but was not at an unexpected level, particularly for a survey of organizations. Prior survey work suggests that organizational survey response rates in the teens or lower are not uncommon. For example, a 2010 mail survey of manufacturers in the West Michigan region obtained a response rate of 16.5 percent, and a nationwide mail survey of economic development organizations conducted the same year had a response rate of 16.4 percent (Watts & Erickcek, 2010; Erickcek, Watts, Robey, Robey, & O’Brian, forthcoming). Both of these surveys included an introduction from an organization known to the respondent and were conducted as part of projects with budgets that allowed for two mailings, which should have helped boost the response rate. Conversely, the survey of CFs was able to conduct only one mailing and was not accompanied by an introduction or endorsement from an organization or individual widely known to CFs, such as a trade organization or a representative of a well-known CF. Still, the response rate for the survey of CFs used in this study was similar to the previously mentioned surveys.

Overall, the evidence suggests that the response rate falls within an acceptable, if not expected, range. A meta-analysis of organizational surveys published in 2005 found an average response rate of 35 percent with a standard deviation of 18.2 percentage points (Baruch & Holton 2008, p. 1149). Although the average is higher than that obtained for the CF survey, the standard suggested by Baruch and Holton (2008) is that response rates should fall within one standard deviation of the mean, which would place
the response rate for the survey of CFs just slightly lower than the suggested minimum standard, 16.8 percent. Given that researchers have documented consistently declining response rates over time (for example, see Dixon & Tucker, 2010; Baruch & Holton, 2008) it seems likely that the expected response rate has only fallen further. If the rate of decline reported by Baruch and Holton (2008) between 2000 and 2005 is assumed to continue through the 2005 to 2010 period then the average response rate of organizational survey response rates could be expected to drop to 33.8 percent. Assuming no change in the standard deviation, the CF survey response rate of 15.7 percent falls within the proposed guideline of one standard deviation from the response rate mean.

It should also be noted that higher organizational response rates are often the product of techniques that have been shown to boost responses but that are not available in all situations. For example, in a survey on CF organizational culture, Keiser (2000) was able to obtain a response rate of greater than 50 percent for a small survey of a geographically-bounded sample of CFs by utilizing multiple techniques designed to boost the response rate. The survey was restricted to CFs in only one state and Keiser (2000) was able to obtain a letter of introduction from a state-level organization that was well known to the participants. Additionally, the research utilized a survey form that had already been field tested by a previous researcher and a follow-up mailing was also used to encourage participation.

This study faced barriers that Keiser did not. Obtaining a high response rate was more difficult because of the larger size and diversity of the survey population, the absence of unique contacts who could provide a letter of introduction or lend additional credibility to the study, and budgetary limitations that prevented follow-up contact with
non-respondents. The result is a response rate that is on the low end of the expected range for organizational surveys.

However, response rate is not nearly as important obtaining a sample that is unbiased and of adequate size to generate estimates with a minimal amount of error. The next section describes the process for determining that a sample is of adequate size. Then, the following section discusses the issue of sample bias.

**Survey Sample Size**

One important aspect of the survey is whether the total number of responses received provides a large enough sample of the overall survey population to provide value estimates that are within an acceptable range of certainty. Although there is no exact standard, it is possible to calculate the sample size needed to achieve a specific alpha or level of acceptable error. Bartlett, Kotrlik, and Higgins (2001) describe how the basic sample size of a survey can be determined if the variance and acceptable level of error are known by using the following equation:

\[ n = \frac{t^2 \times s^2}{d^2} \]  

(3.1)

Where variables represent the following values described below:

- \( n \) = basic sample size
- \( t \) = the \( t \) value for the selected alpha in each tail of the distribution
- \( s \) = the estimate of standard deviation in the population
- \( d \) = the acceptable margin of error for the mean being estimated

With the equation established, known or estimated values were then used to calculate the basic sample size needed for the survey. For \( 't' \) a value of 1.96 was selected,
which represents the value equal to an alpha of 0.05, where 0.025 is in each tail of an
expected distribution curve. This level was chosen because an alpha of 0.05 is a typically
used minimum academic standard in the social sciences. To determine ‘s’ it was
necessary to estimate the standard deviation of the population; however, because the
survey uses multiple questions and there was not previously existing data available from
which to derive a calculation, a scaled variable variance estimation technique was used.
Since the majority of the questions on the survey were designed to collect either scaled
continuous ratings or categorical responses, the variance estimation calculation described
by Bartlett, Kotrlik, and Higgins (2001) was deemed appropriate.

The equation used is as follows (3.2). The number of points on the scale, 6, was
selected because it represents the size of the largest scale used in the survey. The number
of standard deviations, 5, represents the estimated area around a theoretical mean that
would capture that would capture more than 95 percent of all responses. Although this
condition is not strictly known, it is useful as an estimate for the purposes of this
calculation as described by Bartlett, Kotrlik, and Higgins (2001).

\[
\frac{\text{Points on the scale}}{\text{Number of standard deviation}} = \frac{6}{5} = 1.20
\]

Finally, the value of ‘d’ is estimated by multiplying the number of points on the
scale by the acceptable margin of error. The acceptable margin of error selected was
0.03, which corresponds with the minimum standard suggested by Krejcie and Morgan
(1970, as cited in Bartlett, Kotrlik, and Higgins, 2001) for academic work using
continuous data. As mentioned previously, there are 6 scale points. Using the
aforementioned values in equation 3.1 results in an initial basic sample size determination of 171, as shown below.

\[
\frac{1.96^2 \times 1.2^2}{(6 \times 0.03)^2} = 170.74
\]  

(3.3)

The basic sample size determination calculated in equation 3.3 exceeds 5 percent of the population \((171 / 887 = 19.2\%)\), which is an excessively large sample; therefore, a finite population correction was used to calculate a more appropriate sample size. Cochran’s (1977, cited in Bartlett, Kotrlik, and Higgins 2001) correction formula is shown below.

\[
n_1 = \frac{n_0}{1 + \frac{n_0}{\text{Population}}}
\]  

(3.4)

\(n_0\) = previously calculated basic sample size  
\(n_1\) = adjusted sample size

Substituting the previously calculated sample size, 171, and the CF population of 888 derived from the Foundation Directory Online (2010) listing into equation 3.4 results in a corrected sample size of 143. The actual number of responses obtained, 139, is nearly equal to the corrected basic sample size estimate of 143. This suggests that the survey sample obtained for this study is approximately in line with the estimated level expected for rigorous academic work. However, because the actual sample is slightly under the size recommended by the equation, it is possible that levels of variance will be just slightly higher than the desired level.

**Possible Sources of Bias**

Survey response rate and sample size may be the most basic measures of survey validity; however, even a large survey response would not automatically guarantee that
the data provided will represent means that are representative of the overall population. Although organizations do not possess the same demographic characteristics that can be associated with response variation in surveys of individuals (such as gender, age, and ethnicity), it is still possible that some CF traits may bias the findings if the sample of CFs is not representative of the composition of the overall population. Survey bias, which Czaja and Blair (2005) describe as occurring when a "measurement tends to be consistently higher or lower than the true population value" (p. 195) can potentially contribute to Type I error, where the null hypothesis is incorrectly rejected, or Type II error, where a null hypothesis is incorrectly maintained.

**Testing for Bias Using Known Characteristics**

One way to address the issue of possible bias is to examine the characteristics of the responding organizations and compare them with the known characteristics of the overall population of CFs. To achieve an unbiased survey sample that is truly representative of the population being sampled, the essential characteristics of the sample should be identical to those of the population from which they were drawn. If specific characteristics of the respondent are thought to influence the answers provided to questions on the survey, then it is particularly important to ensure that the distribution of these characteristics among respondents is representative of the overall population; otherwise, the estimates derived from the sample will be biased and will not reflect the true proportion of responses that would be given by the population.

Of course, not all the characteristics of an individual or organization that is being surveyed are known to the researcher. The remainder of this section describes techniques used to test for possible sources of bias using characteristics that were known for the
surveyed CFs, as well as the overall population of CFs. Of course, not all possible sources of bias are known or testable; therefore, additional sources of bias that may influence the results of the survey are discussed in the following section.

The first trait examined as a possible source of bias was geography. CF locations are one of the few traits already known for the entire population, and differences in geographic location could possibly be associated with response variation related to cultural differences, socioeconomic differences, and even familiarity with the researcher’s sponsoring organization (if CFs in the Midwest are more likely to recognize Western Michigan University, for example). To test for possible bias, all CFs were categorized into one of nine U.S. Census Bureau Divisions (Census Divisions) to allow for a comparison of the geographic distribution. Although many geographic classifications are possible, the U.S. Census Bureau definitions were selected because they are well established and are described by the Census Bureau (n.d.) as being relatively homogenous when the regional boundary definitions were established. A list defining the composition of each U.S. Census Bureau division is provided in the appendix.

The table below shows the distribution of survey respondents and non-respondents by Census Division (Table 1). An analysis of the data using Stata, a well-known statistical software package, indicates that there is not a statistically significant difference between the expected distribution and the observed distribution of the survey respondents and non-respondents, based on a chi-square statistic of 7.02 with 8 degrees of freedom, which results in a p-value of 0.52. This suggests that geographic differences in the location of respondents are not likely to be a source of bias in the survey results. It
should be noted that one response, from a CF in Puerto Rico, was excluded because the Census Division definitions exclude U.S. territories.

Table 1

<table>
<thead>
<tr>
<th>Census Division</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrespondent</td>
<td>29</td>
<td>63</td>
<td>246</td>
<td>110</td>
<td>117</td>
<td>24</td>
<td>40</td>
<td>32</td>
<td>88</td>
</tr>
<tr>
<td>Respondent</td>
<td>6</td>
<td>10</td>
<td>58</td>
<td>17</td>
<td>19</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>10</td>
</tr>
</tbody>
</table>

The second trait of concern as a possible source of bias was fund endowment size. If nonresponse was in some way associated with a CF’s asset size, it could bias the sample by producing results that are only relevant to larger or smaller CFs instead of the overall population of CFs. The issue of endowment size response bias is also of particular importance, since endowment size is examined for a relationship with measures of evaluation usage later in the analysis section.

To test for a bias in the sample related to asset size, the distribution of the sample of responding CFs and the overall population of CFs was examined. If the CFs that responded represent a similar mix of assets as the overall population, it provides evidence that there is not a bias in the sample. However, if the respondent group contains a different distribution of large- and small-endowment CFs, then it suggests that the results derived from the sample of respondents may be biased by an overweighting of responses from CFs of a certain endowment size.
As a check, the Kolmogorov-Smirnov (K-S) test on the distribution of CF endowment size of the CF population and respondent sample was conducted using Stata. The K-S test is a well established statistical method for examining the goodness-of-fit of two known continuous distributions that offers a more precise estimate than chi-square analysis or simple comparison of histograms (Massey, 1951). An advantage of the K-S test is that it is more precise than chi-square analysis and can produce a robust estimate from small sample sizes. Additionally, the K-S test is nonparametric, which is to say that the K-S test does not require the assumption of a normal distribution.

The null hypothesis of the K-S test is that the sample and the population have the same probability distribution. Based on a K-S test statistic of 0.2414 and a p-value of <0.01 the null hypothesis was rejected by the test, which indicates that the distributions of the respondent CF endowments and the nonrespondent CF endowments are not equal. This suggests that there may be response bias in terms of the relative endowment size of CFs that responded to the survey.

As shown in Table 2, using data from the Foundation Center Directory, the average endowment size of responding CFs is higher than for nonresponding CFs. When self-reported data on the respondents taken from the CF survey is used, the average is even higher; however, this may simply represent rounding up by respondents. It is also possible that an overall increase in CF endowment assets may have occurred between the time of publication of the Foundation Directory and the time when the survey was conducted.
Table 2

CF Endowment Descriptive Statistics

<table>
<thead>
<tr>
<th>Data Source</th>
<th>CF Group</th>
<th>N</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF Directory</td>
<td>Non-respondents</td>
<td>609</td>
<td>62,100,000</td>
</tr>
<tr>
<td>CF Directory</td>
<td>Respondents</td>
<td>124</td>
<td>77,700,000</td>
</tr>
<tr>
<td>Survey</td>
<td>Respondents</td>
<td>136</td>
<td>87,579,670</td>
</tr>
</tbody>
</table>

The evidence on endowment size suggests that there could be a bias in the survey respondents. As reflected by the average endowment size for respondent and nonrespondent CFs, it appears that larger CFs may have replied at a slightly higher rate than smaller CFs. The possible upward bias in the sample could impact the results, particularly since one hypothesis of this study is that CFs with larger endowments may be expected to have a higher predilection toward evaluation than CFs with fewer endowment assets to work with. Indeed, the fact that CFs with larger endowments appear to have responded to the survey at a somewhat higher rate than CFs with smaller endowments may support this premise if an interest in evaluation also extends to an interest in participating in evaluation surveys.

Another way of testing for a difference in endowment size between survey respondents and non-respondents is by using a simple means test, also known as a t-test, to determine whether the difference between the average endowment sizes of the CFs are statistically different. Using Stata to conduct a t-test of means between the two groups produces a t-stat of -0.73, which indicates that there is not a statistically significant difference between the mean endowment size of the respondents and the mean endowment of the non-respondents, based on a p-value of greater than 0.05. The t-test
results suggest that the difference between the respondent and non-respondent group is likely small in practical terms. However, the K-S test remains the superior test of distribution, since it is more sensitive than most other alternatives (Lilliefors, 1967). As a result, the results of this study are carefully interpreted since a possible source of bias related to the endowments held by participating CFs has been identified.

**Other Possible Sources of Bias**

In addition to identifiable and testable source of bias discussed earlier, there are several other possible sources of bias that should be discussed, even though they are not readily testable through empirical means. Fortunately, some of these possible sources of bias are unlikely to be a concern because of the nature of the survey. For example, recall bias, which occurs when respondents are asked to answer questions about previous situations or events, was not considered a major concern, since the survey questions asked primarily about current conditions or recent actions at the CF and did not require the respondent to remember significant past details about the organization. Additionally, the issue of bias arising from sampling frame problems was not examined, since the survey was administered to the entire nationwide population of CFs, as defined by the listing of CFs in the Foundation Center (2010) Foundation Directory Online.

Of course, the fact that the entire population of CFs received the survey, yet only a small portion choose to respond does bring up the issue of response bias related to factors outside the realm of the previously examined sources of bias. In general, there are several possible reasons that can influence whether or not an organization replies to a survey that may not be related to other, observable characteristics of the organization. Factors that influence organizational response include the expertise or authority of the
individual that received the survey in the mail, the interest of the individual that received the survey letter in the topic of the survey, and the degree to which the organization was active in the topic being surveyed (Tomaskovic-Devey, Leiter, & Thompson, 1994). Unfortunately, the presence of these sources of bias is not easy to measure.

One issue is that the individual who received the survey letter in the mail may have played a large role in determining whether or not the survey was completed and returned on behalf of the CF. It was originally desired for the mailing to direct the survey to an individual in a leadership role, whom would presumably have the knowledge and authority to answer questions about the overall organization; however, it is a known issue that the mailing did not have a consistent reach. One issue was that the mailing list obtained from the Foundation Directory Online (2010) did not provide a contact person for every listed CF; conversely, even when a contact was listed, the title and number of persons listed was not consistent. A large portion, 64.9 percent, of the mailings were able to be sent to CF leaders (defined as those with a title of President, Executive Director, or CEO); however, 20.7 percent were addressed to individuals with other non-leadership titles or no listed title, and 14.4 percent of the surveys were sent to CFs where it was not possible to address the letter to a specific contact person.

In addition to the inability to consistently direct the surveys to contacts at a specific leadership level within the organization, it is also highly likely that other internal factors played a role in both the routing of the survey within the organization and whether the individual who finally received the survey choose to respond. For example, it is likely that some portion of the mail addressed to CF leaders was initially opened and reviewed by another staff member, such as an administrative assistant, that may have
screened or redirected the mailing. Ultimately, the determination as to whether or not a survey was completed or returned was likely related to whether or not the mailing reached an appropriate individual within the organization; however, identifying the proper individual and the best route to reach them was simply not possible to determine based on the limited available information in the mailing list.

Of course, obtaining a response was not only related to reaching an individual with the capability and authority to answer a survey, but also on whether or not the specific individual representative that was reached was interested in the topic of the survey. As Czaja and Blair (2005) note, "in a mail survey, nonresponse is more strongly related to interest in the topic than in telephone surveys where the decision to participate or not is made before knowing very much about the survey content" (p. 39). In short, because potential respondents were able to quickly surmise the length, difficulty, and topic of the mail survey, it is likely that individuals with either a personal interest in evaluation or that represented CFs that are active evaluators were more apt to respond. The result is that the estimates derived from this survey may face a positive skew, where responses on measures of evaluation practice and use are more likely to be affirmative than they would be in the overall population.

Unfortunately, the presence or magnitude of a bias associated with higher rates of response from CFs that practice or have a greater interest in evaluation is almost impossible to either verify or quantify. If it were true that CFs with larger endowment sizes are more likely to be evaluative than other CFs, then the fact that the previous examination of the distribution of respondents using the K-S test revealed a small difference in distribution may be highly relevant. In short, a somewhat greater response
rate by larger-endowment CFs may be an indicator that evaluative CFs have indeed responded at a higher rate; therefore, the results of this study, particularly the estimates of evaluation practice and usage, must be interpreted with the assumption that there exists some degree of bias.

The results of this study, as discussed in greater detail in the following chapters, likely overestimate the level of evaluation practice and use among the overall population of CFs. On one hand, this should not be a major issue for the general conclusions of the study, which suggest that evaluation practice and use are still at somewhat low levels. However, specific estimate levels could be assumed to potentially contain a higher margin of error than calculated based on the standard assumption of a representative sample. The issue of bias is discussed further in Chapter V in the context of the overall conclusions of the study.

**Survey Data Collection and Analysis**

All data were collected from the returned survey forms and was entered by the author into a Microsoft Excel database, which was then converted into a format compatible with statistical analysis software. The software package Stata was used to generate all quantitative results. All analysis results are detailed in Chapter IV. A sample copy of the survey instrument is included in Appendix A.

**In Depth Data Collection from Select Community Foundations**

In addition to completing the mail survey, a small subsample of the CFs that participated in the mail sample also participated in a more in-depth phone interview. The purpose of these follow-up sessions was to collect more detailed information on why CFs choose to use or not use evaluation, as well as to provide additional evidence about the
general claims examined through questions in the mail survey. This section gives a brief overview of the qualitative research methodology employed in this study. A more detailed description of the process is described in Chapter V.

**Selection of Participants and Bias**

The CFs participating in the interviews represent a non-random sample, since the interviewees opted in by indicating on the mail survey form they returned that they would be interested in completing an additional interview session. All CFs that opted-in and provided valid contact information were contacted for follow up interviews, although not all elected to ultimately participate.

It should also be noted that the sample of CFs that participated in the follow-up sessions was likely to be biased toward higher-performing CFs and CFs that are more likely to use and have an active interest in evaluation. Organizations that illustrate poor evaluation practices might be expected to shy away from sharing what could be embarrassing stories, while CFs with little evaluation experience may feel they have nothing to share. As a result, the participants may reasonably be expected to represent some of the better and more dynamic CFs in the overall survey sample, which are not representative of a typical CF. However, since the purpose of the case studies was show examples of how CFs can use evaluation to provide value to their communities and donors, a representative sample was not required.

**Qualitative Methodology**

Each CF that opted into the case study portion of the research was contacted to participate in an open-ended, semi-structured interview. Phone calls were scheduled at a convenient time for the participant over a two-week period in February 2011. Potential
participants were also given the option of providing comments or engaging in an interview discussion via email if a convenient time could not be set, however, no participants selected the latter option. The primary topic of discussion was the CF’s use of evaluation, although additional questions about the organization’s size, donor base, growth, and mission were also asked.

The purpose of the qualitative research portion of this study was twofold: triangulation of the results of the quantitative analysis of the mail survey results and development of narrative description and theory related to both why and how CFs practice and use evaluation. All analysis was conducted by the author. The results of the qualitative analysis are discussed further in Chapter V.

**Analysis Approach and General Hypothesis**

After the data were collected, several analysis steps were employed to examine the data in the context of addressing several general hypothesis regarding CF practice and usage of evaluation. Based on the review of literature, the following basic hypotheses were generated:

1. Evaluation practice and usage is still not widespread among CFs.
2. Overall, evaluation practice and usage is increasing.
3. The acceptance of various evaluation roles differs across CFs.
4. The size of a CF plays a role in whether and how evaluation is utilized by the organization.

The first three general hypotheses were examined by analyzing the results of survey questions. Simple descriptive statistics were used to illustrate the responses of the CF sample group to each of the survey questions. Results from the simple statistical analysis
are discussed in the context of the degree that each represents a measure of evaluation practice or usage by CFs.

To address the fourth hypothesis, statistical testing was conducted to check for differences between larger and smaller CFs. First, the respondent CFs were categorized into large and non-large size groups based on the two main measures of CF size obtained from the survey: endowment and population of the service area. The categorical data are then analyzed using chi-square analysis to test for differences in the distribution of responses on questions measuring evaluation practice and usage between those CFs categorized as large and other CFs. The analysis process and results derived from the survey are detailed in Chapter IV.

Finally, the general hypotheses were also examined during the analysis of the qualitative data collected through the follow-up interview sessions. The information provided by CF participants in the follow-up data was used as a source of additional evidence on the findings derived from the analysis of the survey data, as well as a way of providing depth and explanation about possible reasons why CFs do or do not evaluate. Further detail and the results of the qualitative analysis are provided in Chapter V.
CHAPTER IV

SURVEY RESULTS

This chapter presents an analysis of the results of the mail survey of CFs described in detail in Chapter III. The major research question that is addressed by this analysis is as follows: to what degree—considering both prevalence and quality of practice—do CFs currently engage in evaluation? To this end, two major sections of analysis and commentary are presented. The first section examines the broad question of the degree to which CFs conduct evaluation, and presents a quantitative analysis of the results of each survey question for the entire spectrum of CFs. Second, the results are broken down by CF characteristics, such as asset size or service area size, to test as to whether there are inherent CF traits that affect the degree to which CFs engage in evaluation.

General Survey Results

The first question of the survey asked respondents to indicate activities for which the foundation had in place formal procedures for evaluating. Respondents were requested to select all responses that apply and were also given the option of providing an open-ended “other” response in addition to the preselected categories, if they felt that their CF had in place evaluation procedures for activities not already listed. The “other” option was selected by 0.7 percent of respondents—equal to only one CF—with an indication that there were formal employee evaluation procedures in place.
As shown in Table 3, the evaluation of applications from individuals seeking grant funding or financial support was the most popular type of evaluation and was selected by a large majority of the responding CFs. It was not surprising that the vast majority of CFs reported having procedures in place to assess grant applications; still, the response rate suggests that as many as one out of every five CFs do not have a formal grant applicant review system in place for organizations. In theory, all CFs should have an interest in reviewing grant applications in some systematic fashion to ensure that, at a minimum, the funds are distributed to individuals who meet some basic requirement of need, interest, residency, or other requirements.

If CFs do not have in place systematic procedures for evaluating grant applications, it suggests that perhaps informal evaluation procedures are being frequently used or that evaluation procedures vary across different types of grants and procedures. Conversely, it may also reflect a rise in restricted funds that are given away based on rigid pre-set guidelines to qualified applicants, as opposed to being granted on a competitive basis or a basis of need.

The activity that the smallest share of CFs reported evaluating was their own internal activities. About one-third of survey respondents indicated that they had self-assessment procedures in place (Table 3). This is somewhat concerning, since the performance of organizations that typically control millions of dollars should be of interest to both donors and the communities they serve. However, as discussed earlier in the literature review section, this finding does confirm previous researcher findings suggesting that CFs are rarely evaluated.
Table 3
In-Place Systematic Procedures to Conduct Evaluations

<table>
<thead>
<tr>
<th>Response</th>
<th>Percent of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of the needs of the community you serve (e.g. local social and economic issues)</td>
<td>44.6</td>
</tr>
<tr>
<td>Evaluation of applications from individuals seeking grant funding or financial support</td>
<td>45.3</td>
</tr>
<tr>
<td>Evaluation of applications for funding from organizations</td>
<td>81.3</td>
</tr>
<tr>
<td>Evaluation of the activities of non-profit organizations funded by your foundation</td>
<td>56.8</td>
</tr>
<tr>
<td>Evaluation of the outcomes of activities funded by your foundation</td>
<td>57.6</td>
</tr>
<tr>
<td>Evaluation of the effectiveness of your foundation's own activities</td>
<td>33.8</td>
</tr>
<tr>
<td>Other</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td><strong>139</strong></td>
</tr>
</tbody>
</table>

The second question on the survey asked respondents about the maximum degree to which CFs encourage or even require evaluation practice from grantees. Respondents were instructed to select only the one description that best fit their organization’s stance on evaluation practice by grantees, although some organizations may have policies that vary depending on factors such as type or size of the grant. As Table 4 illustrates, a plurality of CFs indicated that grantees are required to have an evaluation component and the majority of respondents indicated that this or a higher level of evaluation was required for grantees. However, the responses also indicate that one-in-four respondents have no formal requirement for any type of grantee evaluation. This does not necessarily mean that the grantees of one-fourth of all CFs go totally unevaluated, since it is possible that program staff may request evaluations based on informal guidelines; however, it does
indicate that a large portion of CFs still do not include evaluation as an integral part of their activities.

Table 4
Level of Encouragement or Requirement for Grantee Evaluation Practice

<table>
<thead>
<tr>
<th>Response</th>
<th>Percent of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is no formal requirement</td>
<td>24.8</td>
</tr>
<tr>
<td>Proposals are assessed based on inclusion of an evaluation component</td>
<td>12.0</td>
</tr>
<tr>
<td>Grantees are required to have an evaluation component</td>
<td>30.1</td>
</tr>
<tr>
<td>Some types of grantees, but not all, are required to incorporate evaluation</td>
<td>18.1</td>
</tr>
<tr>
<td>The foundation conducts its own evaluations of grantees and funded programs</td>
<td>14.3</td>
</tr>
<tr>
<td>The foundation uses an outside consultant to evaluate grantees and funded programs</td>
<td>0.8</td>
</tr>
<tr>
<td>N</td>
<td>133</td>
</tr>
</tbody>
</table>

It should also be noted that even among CFs that do encourage or require evaluation for grantees, a relatively small portion actually directly engage in evaluation activities themselves. According to the survey, only 14.3 percent of CFs actually conduct their own evaluations of grantees or grant-funded projects, and almost no CFs hire outside consulting firms to conduct grantee and funded-project evaluations on their behalf. This may be a function of limited resources, wherein CFs prefer to play the role of building the capacity of grantees to conduct their own evaluations, since multiple respondents indicated in the open-ended final question of the survey that either financial or staff resources were a factor in restricting their own evaluation activity. However, the fact that most CFs do not regularly play the role of evaluator, also suggests that CFs must
place a great deal of trust in evaluations conducted by grantees or other outside parties when conducting oversight.

The third survey question asked CFs about how evaluation results are used by the organization. Respondents were presented with 11 possible evaluation uses and asked to rate them using a four-point scale, which ranged from a value of 1 (indicating that results are not used for that purpose) to 4 (indicating that evaluation results are almost always used for that purpose.) Survey respondents were also given the option of describing another activity that they use evaluation for if it was not otherwise listed; six of the respondents wrote in the following activities:

- Formative training programs
- Continuous learning
- Real-time feedback during the course of the grant to improve performance of the grantee and enhance opportunity to reach outcomes
- Effectiveness of approach and impact of major initiatives
- To show donors the effectiveness of their grant
- To ensure grants comply with national standards for CFs

Table 5 shows the percentage response of CFs on the four-point scale for each of the 11 major evaluation purposes they were asked to consider. The evaluation usage purpose that got the highest overall response was “to approve or select grants for financial support” which was a cited as having a frequent or almost always level of usage by 72.3 percent of respondents, while only 2.9 percent indicated that they do not ever use evaluation for this purpose. This finding only further confirms that the assessment of
grant applications may be the one consistent area of evaluation for CFs, which was also highlighted in the first survey question.

Table 5

Ratings of Evaluation Usage by CFs for Selected Purposes

<table>
<thead>
<tr>
<th>Purpose of Evaluation Usage</th>
<th>Percent Responding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
</tr>
<tr>
<td>To approve or select grant applicants for financial support (n=137)</td>
<td>2.9</td>
</tr>
<tr>
<td>To identify the most pressing problems in the community (n=137)</td>
<td>8.8</td>
</tr>
<tr>
<td>To assess our own operations (n=138)</td>
<td>13.8</td>
</tr>
<tr>
<td>To support fundraising efforts or to show results to donors (n=136)</td>
<td>15.4</td>
</tr>
<tr>
<td>To provide board members with information for decision making (n=138)</td>
<td>2.90</td>
</tr>
<tr>
<td>To help local non-profit organizations improve their operations (n=137)</td>
<td>21.2</td>
</tr>
<tr>
<td>To publicize our foundations achievements to the community (n=138)</td>
<td>12.3</td>
</tr>
<tr>
<td>To fulfill the requirements to receive funding from other foundations or government agencies (n=136)</td>
<td>38.2</td>
</tr>
<tr>
<td>To ensure that organizations/programs funded by the foundation meet legal requirements (n=136)</td>
<td>10.3</td>
</tr>
<tr>
<td>To generate information on the validity of a type of program or theory (e.g. to demonstrate effectiveness of a type of poverty alleviation program) (n=136)</td>
<td>27.2</td>
</tr>
<tr>
<td>To select programs or funding opportunities that should receive additional funding (n=138)</td>
<td>15.2</td>
</tr>
</tbody>
</table>
Other purposes that evaluation results often are used for included ensuring that organizations and programs funded by the foundation met legal requirements, which was listed as being a frequent or almost always usage by 69.1 percent of responding CFs, and providing board members with information for decision making, which was listed as being a frequent or almost always usage by 67.4 percent of respondents. At the other end of the spectrum, the lowest-rated uses for evaluation seem to suggest that many CFs are not concerned with (or perhaps even resistant to) the use of evaluation for purposes outside of the CF organization. For example, nearly two-thirds of respondents indicated that the usage of evaluation to assist in the improvement of operations at local non-profit organizations was either minimal or did not occur at all. CFs also do not appear to have a strong interest in using evaluation to understand the effectiveness of broad approaches to community issues, as is indicated by the low ratings for usage of evaluation “to generate information on the validity of a type of program or theory.

The fourth survey question deals with whether CFs use or promote specific standards or guidelines for evaluation. The purpose of this question was to begin addressing the concept of quality, since in addition to the fact that evaluation is known to not be a universal practice among CFs, the relative rigor of the evaluations that are conducted remains virtually unknown. Respondents were provided with five statements describing possible levels of guideline promotion that they might do and were asked to select the one that best fit. An option of providing an open-ended “other” answer was also provided; however, only two respondents choose to provide an alternate response. One CF indicated that grantees were required to work with a specific evaluation
consultant, which was paid for by the CF; the other respondent indicated that their only specific guideline was requiring the creation of a logic model.

As Table 6 illustrates, the majority of CFs do not have specific guidelines either for their own evaluation practices or that are recommended to grantee organizations for conducting their own evaluations. Nearly half of respondents either do not specify standards or guidelines, or they have no requirements for grantee evaluation in the first place and therefore cannot specify guidelines or requirements. Among CFs that do have requirements or recommendations for how grantee evaluations are done, only 2.9 percent utilize standards that were created by an outside organization.

Table 6
Recommendation or Promotion of Evaluation Guidelines by CFs

<table>
<thead>
<tr>
<th>Description of Approach</th>
<th>Percent of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grantees are not required to be evaluated and therefore no guidelines are provided</td>
<td>16.1</td>
</tr>
<tr>
<td>Grantees and programs are required to be evaluated, but we do not specify any specific standards or guidelines</td>
<td>29.9</td>
</tr>
<tr>
<td>Sometimes specific guidelines or checklists are recommended or required, but it depends on the situation</td>
<td>24.1</td>
</tr>
<tr>
<td>We recommend or require the use of a standard set of guidelines or a checklist that has been created by an outside organization (e.g. Key Evaluation Checklist, Joint Committee on Standards, the CIPP model, etc.)</td>
<td>2.9</td>
</tr>
<tr>
<td>We recommend or require the use of evaluation guidelines or standards that were specially developed by our foundation</td>
<td>25.6</td>
</tr>
<tr>
<td>Other</td>
<td>1.5</td>
</tr>
<tr>
<td>N</td>
<td>137</td>
</tr>
</tbody>
</table>
Although models such as Scriven's (2007) Key Evaluation Checklist or Stufflebeam's (2007) CIPP model are widely recognized in the fields of educational evaluation and external program evaluation, this survey result indicates that these and other models have not yet found recognition among professional CF staff. Instead, CFs were far more likely to report that they had created their own set of standards or guidelines, or that recommendations regarding how grantee evaluations should be conducted were only made under certain circumstances.

The lack of recognition by CFs for formal models developed by evaluation theorists or other organizations does not, on its own, necessarily mean that CFs are failing to promote high-quality evaluation practice. As was shown in Table 4, approximately one-quarter of CFs consistently recommend or require their own guidelines, and nearly one-quarter of CFs at least makes some sort of recommendation concerning evaluation practice when warranted by the project or grantee. To address this, the next question in the survey asked respondents to rate the importance of several broad factors in terms of how often they are present in evaluations conducted by or for the organization. Each of the listed factor descriptions was based loosely on the four major areas covered in The Program Evaluation Standards 2nd Edition: utility standards, feasibility standards, propriety standards, and standards related to accuracy (Joint Committee on Standards for Educational Evaluation, 1994).

Although the Joint Committee standards were originally drafted with a primary focus on the evaluation of educational activities, the four main categories of standards can be applicable to many types of evaluation, since they deal with broad concepts of quality and usability that are applicable across fields. Indeed, these standards have been
adopted for use in fields such as public health evaluation (Centers for Disease Control, n.d.), international development projects (Asian Development Bank, n.d.), and even both the areas of community development and philanthropy (Stufflebeam, 2004). As such, high ratings by survey respondents for the four main classifications of the Joint Committee's evaluation standards could be taken as evidence that CFs are creating or recognizing the basic elements necessary for conducting quality evaluations of grantee and program activities.

Table 7 shows the percentage of respondents rating the degree to which their CF's evaluation practices correspond with a set of factors that describe each of the four main standards put forth by the Joint Committee (1994). The categories of standards that are listed in the table, which were not revealed to survey takers as being standards of the Joint Committee, were described to respondents in the following manner.

- The utility standard was described as “utility of the evaluation, meaning that all key stakeholders have been identified and those conducting the evaluation are credible.”

- The feasibility standard was described as “feasibility, meaning that the evaluation is non-disruptive, sensitive to political issues, and cost-effective.”

- The propriety standard was described as “meets basic standards, meaning that individual rights are respected, strengths and weaknesses are fairly reported, findings are fully disclosed, and conflicts of interest are avoided.”

- The accuracy standard was described as “accuracy is monitored, meaning that grantees’ activities are documented, information is collected in a systematic manner, appropriate analysis techniques are utilized, conclusions are justified by data, and reporting is performed in an impartial manner.”

For a complete listing of how this question was presented to the surveyed CFs, see the copy of the survey form provided in Appendix A.
Table 7
Percent of CF Respondents by Reported Rate of Use of Inclusion of Related Evaluation Practices

<table>
<thead>
<tr>
<th>Category of Basic Joint Committee Standard</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Almost always</th>
<th>Don't know</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility standards</td>
<td>15.04</td>
<td>16.54</td>
<td>22.56</td>
<td>28.57</td>
<td>17.29</td>
<td>133</td>
</tr>
<tr>
<td>Feasibility standards</td>
<td>18.80</td>
<td>11.28</td>
<td>20.30</td>
<td>30.08</td>
<td>19.55</td>
<td>133</td>
</tr>
<tr>
<td>Propriety standards</td>
<td>12.69</td>
<td>2.24</td>
<td>22.39</td>
<td>48.51</td>
<td>14.18</td>
<td>134</td>
</tr>
<tr>
<td>Accuracy standards</td>
<td>12.69</td>
<td>8.21</td>
<td>33.58</td>
<td>29.85</td>
<td>15.67</td>
<td>134</td>
</tr>
</tbody>
</table>

The standards that CFs most widely adhered to are the propriety standards, which 48.5 percent of respondents indicated were “almost always” considered in their evaluation activities. The other three categories of standards—utility, feasibility, and accuracy—were reported as having very similar levels of usage by the responding CFs. In all cases, slightly more than half of CFs indicated that they considered factors related to these standards either sometimes or almost always when conducting an evaluation. Still, a sizable portion of CFs indicated that factors related to the standard were either “never” or “rarely” present when conducting an evaluation.

In addition to capturing measures related to the level and quality of evaluation being conducted, the survey also sought to identify the relative importance of evaluation-related activities. Respondents were asked to rate the importance of eight major activities using a four-point scale. Table 8 shows the percentage of CFs selecting each response for each of the eight selected activities.
Table 8
Percent of CFs Rating the Importance of Select Organizational Activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Not important</th>
<th>Somewhat important</th>
<th>Very important</th>
<th>Essential</th>
<th>Not applicable</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing endowment funds</td>
<td>0.72</td>
<td>2.17</td>
<td>5.07</td>
<td>90.58</td>
<td>1.45</td>
<td>138</td>
</tr>
<tr>
<td>Cultivating new donors</td>
<td>0.72</td>
<td>3.62</td>
<td>13.04</td>
<td>82.61</td>
<td>0</td>
<td>138</td>
</tr>
<tr>
<td>Assessing community needs and identifying the most pressing problems</td>
<td>0</td>
<td>19.57</td>
<td>43.48</td>
<td>36.96</td>
<td>0</td>
<td>138</td>
</tr>
<tr>
<td>Selecting grantees that will have the greatest impact</td>
<td>0</td>
<td>9.49</td>
<td>45.99</td>
<td>44.53</td>
<td>0</td>
<td>137</td>
</tr>
<tr>
<td>Acting as a resource for community information</td>
<td>2.17</td>
<td>13.77</td>
<td>42.75</td>
<td>41.30</td>
<td>0</td>
<td>138</td>
</tr>
<tr>
<td>Supporting the causes that are most important to major donors</td>
<td>3.65</td>
<td>24.82</td>
<td>39.42</td>
<td>29.93</td>
<td>2.19</td>
<td>137</td>
</tr>
<tr>
<td>Ensuring that charitable funds are used by grantees in an efficient manner</td>
<td>0</td>
<td>8.70</td>
<td>32.61</td>
<td>58.70</td>
<td>0</td>
<td>138</td>
</tr>
<tr>
<td>Determining the impact of programs on the local community</td>
<td>1.45</td>
<td>13.77</td>
<td>54.35</td>
<td>30.43</td>
<td>0</td>
<td>138</td>
</tr>
</tbody>
</table>

The activities most frequently rated as being “essential” were those related to maintaining and expanding the financial capacity of CFs: managing endowment funds and cultivating new donors. This is not entirely surprising, given that endowments and donors are present concerns in CFs of every size and age; however, it could lend some
support to the critics who suggest that some CFs are little more than “charitable bankers” (Carson, 2002). If nothing else, this finding does suggest that CFs tend to be more concerned with financial stability than with determining the impact of their programs on the community, an activity that received a lower rating. Of course, it is impossible to tell if this situation indicates a permanent focus of CFs, or if concerns with financial survival merely reflect recent economic times, when donations and endowment growth were both likely to be weak or declining because of the 2007 to 2009 national economic recession.

The activity that received the lowest ratings overall was “supporting the causes that are most important to major donors,” which was rated as “not important” by 3.7 percent of CFs and “essential” by only 29.9 percent of CFs, which was lower than any other activity. This could suggest that CFs feel less beholden to major individual or family donors than other types of foundations, or it may simply be that some CFs do not have any donors that they consider major and instead rely on frequent donations from a variety of community members and businesses.

All of the activities listed in question six of the survey received high ratings: 69 percent or more of responding CFs gave a rating of “very important” or “essential” to all activities. The activities that could be considered to be associated with evaluation usage and practice, assessing community needs and determining the impact of programs, were also deemed as being “very important” or “essential” by a large proportion of CFs: 80.4 percent and 84.8 percent, respectively. Clearly, most CFs agree that these activities are important, however, in a relative sense these activities do fall significantly below donor cultivation and endowment fund management activities in terms of an overall rating of importance by CFs.
The next two questions in the survey addressed the issue of CF staff capabilities and credentials in the practice of evaluation. Question seven asked respondents to indicate if any CF staff members hold formal college degrees in evaluation or assessment, while question eight asked whether any CF staff had completed any short-term training, such as individual courses or seminars, which provided formal training on how to conduct evaluations. These queries also address the issue of evaluation quality in CFs by determining the degree to which CFs have sought out or cultivated evaluation skills in organizational staff. Although having staff that is trained in evaluation practice is neither a prerequisite for nor guarantee of the capacity to conduct high-quality evaluations, the presence of trained staff would suggest that CFs are showing an interest in evaluation as a workplace skill. CFs that invest in staff trained in evaluation, either through hiring or by providing or encouraging existing staff to seek training, are demonstrating an investment in good evaluation practice that goes beyond a simple expression of interest in conducting more or higher-quality evaluations, since cultivating formal skills training usually require a financial investment by the staff member and the CF.

As shown in Table 9, while only a small percentage of CFs employ staff persons that hold a college degree in evaluation or assessment, just over half have one or more staff members that have completed some sort of formal class or seminar on how to conduct evaluations. This is encouraging for those who support a greater emphasis on evaluation in the CF world, since it suggests that many CFs are placing at least some value on staff training in evaluation, particularly if the organization pays the costs of the employee attending the seminar or course on evaluation.
### Table 9

**Questions on CF Staff Training in Evaluation Practice**

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
<th>Don't know</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do any members of CF staff hold college degrees in evaluation or assessment?</td>
<td>9.42</td>
<td>85.51</td>
<td>5.07</td>
<td>138</td>
</tr>
<tr>
<td>Have any members of CF staff attended formal training, such as seminars or college courses, on how to conduct evaluation?</td>
<td>55.80</td>
<td>38.41</td>
<td>5.80</td>
<td>138</td>
</tr>
</tbody>
</table>

The small share of CFs that reported employing staff members that hold college degrees in evaluation is not surprising and most likely reflects the limited availability of college programs in evaluation, particularly at the undergraduate level. The American Evaluation Association (n.d.) list of educational programs reports only 44 U.S. colleges and universities as having graduate degree programs either majoring in or offering a concentration in evaluation. At the undergraduate level the organization lists no programs offering a major or minor in either evaluation or assessment. Although the American Evaluation Association list does not constitute a comprehensive survey of collegiate programs or course offerings in evaluation, it does clearly indicate that the opportunity to gain a specialized degree in the evaluation field is limited compared to other areas of study.

The next three questions address possible uses of evaluation results, such as whether evaluation results are used in cultivating donors or if grantee evaluations have resulted in tangible actions, such as increasing or stopping funding for a grantee program or project. As Table 10 illustrates, a small majority of CFs indicated that they promote evaluation activities to donors; however, CFs were roughly split on the use of evaluation
results in tangible grantee funding decisions. A plurality of respondents, 47.1 percent, reported that their CF had cut funding to a grantee because of poor evaluation results and 42.8 percent reported that they had not, while a sizable 10.14 percent indicated that they did not know. On the question of whether funding was ever expanded for a grantee with strong evaluation results, the results were nearly split with 44.9 percent of CFs responding in the affirmative, 45.6 percent responding no, and 9.6 percent indicating that they did not know.

Table 10

Questions on CF Evaluation Promotion and Usage

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
<th>Don't know</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you ever promoted your evaluation or needs assessment activities to donors?</td>
<td>54.35</td>
<td>43.48</td>
<td>2.17</td>
<td>138</td>
</tr>
<tr>
<td>Have you ever cut funding to a grantee because of poor evaluation results?</td>
<td>47.10</td>
<td>42.75</td>
<td>10.14</td>
<td>138</td>
</tr>
<tr>
<td>Have you ever expanded funding to a grantee because of strong evaluation results?</td>
<td>44.85</td>
<td>45.59</td>
<td>9.56</td>
<td>136</td>
</tr>
</tbody>
</table>

The results of the three survey questions on usage and promotion are not surprising and confirm the results shown in Table 7, which indicated that approximately 50 percent of CFs use evaluation either “minimally” or “not at all” for making decisions on funding. Although it does appear that a slightly larger percentage of CFs may use evaluation when courting donors, it is also clear that a large proportion of CFs are not utilizing evaluations and evaluation results for tangible operational purposes. This is true even when the portion of CFs that do not regularly conduct evaluations is taken into account. Only 16 percent of CFs reported having no requirements at all for grantee
evaluation (Table 6) and more than 80 percent indicated at least some level of usage of evaluation for approving grantee applications, selecting grantees for additional funding, and demonstrating results to donors (Table 5). This suggests that while CFs may see evaluations as having value for these activities, they have a more difficult time documenting instances where evaluation has actually played a direct role in causing an organizational action, such as removing or increasing funding.

Finally, survey respondents were also asked to provide basic information on the size of their CF's endowment and the population size of the community served by the organization. The purpose of these final questions is twofold: to provide a basic measure of the survey sample and to provide a basis for determining if differences in evaluation practice are tied to factors of organizational size. In the next section, survey results are examined across categories of CFs to identify if organizational resources or service areas influence evaluation practice or usage.

As shown in Table 11, the endowment and population sizes of CFs that responded to the survey were widely varied. The endowments of this sample of CFs range in size from $190,000 to $1.6 billion, with an average size of $87.58 million and a median of $19.5 million. It should be noted that the large difference between the mean and median endowment sizes reflects the fact that the sample is comprised of a large number of small CFs and a small number of CFs with extremely large endowments at the other end of the spectrum; in short, there are outliers with very large endowments within the sample. In terms of population size, the communities served by these CFs range in size from a tiny community of only 500 to a major metropolitan region of approximately 10 million
residents. Overall, the survey sample represents a wide diversity of CFs as measured by the organizations’ endowments and service areas.

Table 11

<table>
<thead>
<tr>
<th>CF Endowment and Service Area Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endowment (millions)</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

Analysis for Differences by CF Characteristics

One line of reasoning that has been used to explain the under-use or non-use of evaluation by CFs has been the issue of resource limitations. Respondents to the survey of CFs conducted for this study frequently mentioned both size and resource limitations as reasons for not conducting more grantee evaluations. To test for differences, the survey sample was first broken down into categories using the two main descriptive variables that are available for measuring the characteristics of the responding CFs: endowment size and the population size of the region.

The existing literature on CFs does not offer an accepted definition of what constitutes a large CF endowment or a large CF service area population. Indeed, standards of organizational size have likely changed over time in conjunction with factors such as stock market fluctuations and general growth in popularity of CFs as charitable organizations. Therefore, an approach based on relative distribution was explored to
categorize CFs as large based on their position relative to other CFs in the survey sample. For each of the two general descriptive questions on the survey, endowment size and service area population, the distribution was analyzed and divided based on percentiles. These results are presented later in this section (Table 12).

Next, the idea of combining these two classifications into a single classification of large CF was explored; however, there does not appear to be a particularly strong relationship between CFs with large endowments and CFs that serve large population areas. This finding is based on a simple correlation value of 0.36 between the dummy variables designating CFs with large endowments and CFs with large population service areas. Therefore, the two categories were kept independent and analyzed separately to test whether there is any merit to the previously stated hypothesis that CFs with large endowments or CFs with large service areas will report a different level of evaluation usage than other CFs.

To test for differentiation between the responses of CFs with larger endowments and larger service area populations, a chi-square test of independence was used to see if there is a relationship between the size category and the categorical responses provided on each question. The chi-square test is a way of determining whether the actual observed distribution of categorical responses statistically differs from the expected distribution of responses. If the distribution of categorical survey question responses was found to be different than expected between the groups being examined, then it can be said that there is a statistically significant difference between the actual distribution of responses and the expected distribution. In other words, if the hypothesis that CF size impacts evaluation practice and usage is correct, the distribution of the CFs categorized
as large should differ from the expected distribution. However, if the distributions are the same, then the chi-square test offers evidence that there is not a difference in the responses of the CFs to survey questions measuring evaluation practice.

For each question item, a chi-square statistic was calculated using Stata. Based on the chi-square statistic and the degrees of freedom, a p-value was then calculated to determine whether or not a statistically significant relationship could be detected between the distribution of responses provided by the large endowment or large service area population CFs and the expected distribution of responses of the overall CF sample. The values were calculated three ways: by using the definition of CFs in the 75th, 90th and 95th percentile groups as the “large” CF group that was compared against the remaining CFs.

Only data from questions number two through number 11 of the survey were used in the analysis. The first question was excluded because survey respondents were asked to select all answers that applied to them instead of being limited to a single rating or nominal response; therefore, the resulting data was not appropriate for use in a chi-square analysis, since an expected distribution could not be determined. Question 12 and question 13 were excluded from the analysis because these items collected the data on endowment size and the population of the service area. The final two questions of the survey involved the collection of optional contact information and qualitative data that would not be appropriate for this analysis.

In some instances, a Fisher’s Exact test was used to confirm the relationship because a basic assumption of chi-square analysis was violated: specifically, the expected number of responses for any one possible outcome was too small because of the size of the sample or the unique nature of the distribution. In these situations an additional
test—the Fisher’s Exact test—was utilized instead of chi-square analysis. The benefit of the Fisher’s exact test is that it can be used to directly calculate a probability estimate value of whether or not a nonrandom relationship exists between categorical responses even in situations where one expected response is very low or even zero (Weisstein, n.d.). In most cases, the results of the Fisher’s Exact test were similar to the results of the chi-square test; however, instances where the results of the Fisher’s Exact test differed and were deemed to be the more appropriate measure are noted in the results shown in Table 12.

The Fisher’s Exact test function in the statistical software package Stata was used to generate a p-value when the basic assumption that an expected number of responses for any answer in any category could fall below four. Because this method of conducting the Fisher’s Exact test calculates a specific probability and does not produce a test statistic, the chi-square statistic is reported in all cases, while the results of the Fisher’s Exact test are used to simply confirm that violations of basic assumptions do not alter the findings.

Table 12

Test of Independence by Endowment and Population Category

<table>
<thead>
<tr>
<th>Question</th>
<th>Item</th>
<th>Endowment</th>
<th>Population</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>75 Pct</td>
<td>90 Pct</td>
<td>95 Pct</td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td>18.91***</td>
<td>12.12**</td>
<td>0.87</td>
</tr>
<tr>
<td>Q3</td>
<td>1</td>
<td>1.63</td>
<td>1.19</td>
<td>0.51</td>
</tr>
<tr>
<td>Q3</td>
<td>2</td>
<td>7.63*</td>
<td>2.77</td>
<td>1.07</td>
</tr>
<tr>
<td>Q3</td>
<td>3</td>
<td>4.67</td>
<td>4.61</td>
<td>5.45</td>
</tr>
<tr>
<td>Q3</td>
<td>4</td>
<td>3.27</td>
<td>4.69</td>
<td>3.89</td>
</tr>
</tbody>
</table>
Table 12 – Continued

<table>
<thead>
<tr>
<th>Question</th>
<th>Item</th>
<th>Endowment</th>
<th></th>
<th></th>
<th>Population</th>
<th></th>
<th></th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>75 Pct</td>
<td>90 Pct</td>
<td>95 Pct</td>
<td>75 Pct</td>
<td>90 Pct</td>
<td>95 Pct</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>5</td>
<td>2.63</td>
<td>5.09</td>
<td>5.54*</td>
<td>0.5</td>
<td>0.83</td>
<td>0.9</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>6</td>
<td>2.69</td>
<td>0.15</td>
<td>0.84</td>
<td>3.86</td>
<td>5.3</td>
<td>12.19</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>7</td>
<td>5.43</td>
<td>0.86</td>
<td>0.47</td>
<td>1.26</td>
<td>0.05</td>
<td>0.47</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>8</td>
<td>1.6</td>
<td>2.53</td>
<td>1.82</td>
<td>3.8</td>
<td>10.33**</td>
<td>9.41**</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>9</td>
<td>0.46</td>
<td>0.72</td>
<td>0.03</td>
<td>0.46</td>
<td>0.75</td>
<td>6.77*</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>10</td>
<td>2.78</td>
<td>1.01</td>
<td>0.41</td>
<td>2.78</td>
<td>0.79</td>
<td>1.99</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>11</td>
<td>3.72</td>
<td>6.53*</td>
<td>5.41</td>
<td>3.72</td>
<td>2.25</td>
<td>0.62</td>
<td>3</td>
</tr>
<tr>
<td>Q3</td>
<td>12</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>3.00#</td>
<td>0.375</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td>15.77***</td>
<td>7.41</td>
<td>4.47</td>
<td>8.28</td>
<td>5.53</td>
<td>2.73</td>
<td>5</td>
</tr>
<tr>
<td>Q5</td>
<td>1</td>
<td>3.07</td>
<td>10.71**</td>
<td>10.48**</td>
<td>2.26</td>
<td>2.07</td>
<td>0.35</td>
<td>4</td>
</tr>
<tr>
<td>Q5</td>
<td>2</td>
<td>5.98</td>
<td>12.49**</td>
<td>10.65**</td>
<td>3.93</td>
<td>4.24</td>
<td>0.82</td>
<td>4</td>
</tr>
<tr>
<td>Q5</td>
<td>3</td>
<td>2.11</td>
<td>7.86*</td>
<td>4.7</td>
<td>3.79</td>
<td>2.56</td>
<td>0.86</td>
<td>4</td>
</tr>
<tr>
<td>Q5</td>
<td>4</td>
<td>4.72</td>
<td>6.84</td>
<td>7.37</td>
<td>8.86*</td>
<td>4.75</td>
<td>1.78</td>
<td>4</td>
</tr>
<tr>
<td>Q6</td>
<td>1</td>
<td>7.55+</td>
<td>11.12*</td>
<td>19.15***</td>
<td>1.18</td>
<td>0.94</td>
<td>1.12</td>
<td>4</td>
</tr>
<tr>
<td>Q6</td>
<td>2</td>
<td>3.07</td>
<td>8.29**</td>
<td>13.28#</td>
<td>3.67</td>
<td>1.06</td>
<td>0.62</td>
<td>3</td>
</tr>
<tr>
<td>Q6</td>
<td>3</td>
<td>2.53</td>
<td>0.38</td>
<td>0.77</td>
<td>3.09</td>
<td>0.5</td>
<td>0.06</td>
<td>2</td>
</tr>
<tr>
<td>Q6</td>
<td>4</td>
<td>2.57</td>
<td>0.34</td>
<td>1.13</td>
<td>5.55</td>
<td>3.15</td>
<td>1.53</td>
<td>2</td>
</tr>
<tr>
<td>Q6</td>
<td>5</td>
<td>2.82</td>
<td>2.5</td>
<td>2.65</td>
<td>1.59</td>
<td>3.1</td>
<td>0.56</td>
<td>3</td>
</tr>
<tr>
<td>Q6</td>
<td>6</td>
<td>6.52</td>
<td>0.87</td>
<td>2.72</td>
<td>7.67</td>
<td>5.65</td>
<td>1.87</td>
<td>4</td>
</tr>
<tr>
<td>Q6</td>
<td>7</td>
<td>2.64</td>
<td>2.35</td>
<td>6.29+</td>
<td>1.79</td>
<td>0.66</td>
<td>4.06</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 12 – Continued

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6</td>
<td>8</td>
<td>0.72</td>
<td>3.27</td>
<td>3.03</td>
<td>4.34</td>
<td>1.94</td>
<td>0.63</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td>2.39</td>
<td>7.75**</td>
<td>12.81***</td>
<td>3.97</td>
<td>8.27**</td>
<td>12.07**</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td>10.72***</td>
<td>11.31***</td>
<td>3.73+</td>
<td>0.61</td>
<td>4.56**</td>
<td>1.67</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td></td>
<td>3.31</td>
<td>0.76</td>
<td>0.38</td>
<td>1.17</td>
<td>1.52</td>
<td>1.32</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q10</td>
<td></td>
<td>5.63*</td>
<td>2.34</td>
<td>2.72</td>
<td>5.13*</td>
<td>6.08**</td>
<td>2.51</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11</td>
<td></td>
<td>1.77</td>
<td>6.32**</td>
<td>5.35*</td>
<td>9.10**</td>
<td>9.81***</td>
<td>8.27**</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***p<0.001 **p<0.05 *p<0.10
+ Fishers Exact p<0.1
#Not significant using Fisher's Exact p>0.10

The first step after conducting the initial analysis of the response distributions was to examine the findings and determine whether there were differences between the three percentile-based definitions of “large” CFs. Ideally, if the hypothesis that larger CFs engage more in evaluation than smaller CFs is correct, it could be expected that there would be significant overlap in the findings. However, if the hypothesized effects are only apparent above a certain size threshold or if the hypothesized effects increase in magnitude with size, it is likely that differences in measures of evaluation usage would be more easily detectable at higher levels.

As shown in Table 12, the analysis reveals relatively few statistically significant differences in response distributions by either endowment size or population size, regardless of the percentile level used to define large CFs. When endowment was used to define CFs at and above the 75th percentile, only two questions showed statistically
significant differences with a p-value <0.05, question 4 and question 8. The number rises when the 90th and 95th percentile endowment size definitions are used to seven and four, respectively, also based on a p-value<0.05. In all cases, there are several other questions that show weaker statistically significant differences in distribution, based on a p-value<0.10; however, even at this level the majority of questions show no statistically significant difference in distribution between larger and smaller CFs by endowment size.

A similar story holds true when CF responses were analyzed using the population of the service area to define size. When CFs with population service area sizes at and above the 75th percentile were categorized as large, a statistically significant difference in answer distribution could only be detected for one question, question 11, based on a p-value<0.05. Use of the 90th percentile as a definition increased the number of questions with statistically significant differences in distribution to six, based on a p-value<0.05. When the 95th percentile definition was used the number of statistically significant differences detected declined to three. Again, using a standard of p-value<0.10 increases the number of statistically significant differences in distribution detected using any percentile level definition of large CF; however, the majority of survey questions still would not show differences in distribution regardless of the definition.

One finding that can be taken from this initial analysis is that there appear to be at least some differences between findings depending on what percentile is used to define large CFs. The comparison of results across three different percentile definitions of large CFs—75th, 90th, and 95th—suggests that the 90th percentile is the most appropriate definition of the three for use in this study. As was shown in Table 12, the 90th percentile definition identified a greater number of survey questions with statistically significant
differences in response distribution compared to the 75th percentile level, and identified most of the same differences that were identified when the 95th percentile level was used. Using the endowment variable, the 90th percentile definition identified three of the four questions identified as statistically significant by the 95th percentile definition based on a p-value<0.05. Of the three questions identified as having a statistically significant difference using the service area population and 95th percentile definition, all of the same questions were also identified by the 90th percentile definition. In short, this suggests that both definitions capture much of the same effect, however, the 90th percentile definition provides a larger sample size and may be more appropriate than the 95th percentile definition, which categorizes only a small number of CFs, 11, as being in the large category.

The 90th percentile definition categorizes as large all those CFs with an endowment size of 180 million and higher when the endowment size variable is used or all CFs serving populations of one million and greater as large when the service area variable is used. Throughout the remainder of this study, these definitions will be used exclusively when comparing the survey responses of large CFs compared to other CFs.

Of course, testing the theory that CFs with greater resources engage in more evaluation practice requires establishing not only that a statistically significant difference exists between larger and smaller organizations, but that the difference fits the theoretical expectation that CFs with larger endowment asset sizes or those serving a bigger community demonstrate higher levels of evaluation interest or practice. If the distribution of responses showed a statistically significant difference, but the larger CFs reported lower levels of evaluation usage it would also offer evidence refuting the core
theory. To determine if the identified differences meet theoretical expectations, the
detailed distribution differences between large CFs at the 90th percentile and above
compared to other CFs is discussed in the next section for each statistically significant
question. Questions that did not yield a statistically significant difference are not
discussed in detail; however, the prevalence of this finding is discussed in greater length
at the end of this chapter.

**Survey Response Differences of CFs with Large Endowments**

This section examines individual survey items that were found to have a
statistically significant difference in distribution between the response distribution of
large CFs and all other CFs. The detailed analysis of each item is necessary to determine
whether the responses fit a pattern that meet the hypothetical expectation that larger CFs
will indicate higher levels of evaluation usage. For each item meeting the standard of
statistical significance, a table comparing the percent of survey respondents that selected
each scale or category response is presented and the difference and its confirmation or
rejection of the hypothesis is discussed.

Although the summary table of the chi-square and Fishers Exact test analysis
notes multiple levels of significance, all survey items examined in this section are
statistically significant based on a p-value<0.05 using the appropriate chi-square or
Fisher’s Exact test. In total, when the 90th percentile endowment definition of large CFs
was used, six survey items showed a statistically significant distributional difference
across groups. The results of the analysis when using the 90th percentile service area size
definition are presented in the following section.
The first question to show a statistically significant difference in answer distribution when divided by endowment size category is survey question number two. The second survey question addressed the degree to which CFs require grantees to practice evaluation. If the theory that CFs with better funding engage in, or support, more evaluation holds true, then it would be expected that CFs with large endowments will more frequently require evaluation than other CFs.

Table 13

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is no formal requirement</td>
<td>25.6</td>
<td>18.8</td>
</tr>
<tr>
<td>Proposals are assessed based on inclusion of an evaluation component</td>
<td>12.8</td>
<td>6.3</td>
</tr>
<tr>
<td>Grantees are required to have an evaluation component</td>
<td>32.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Some types of grantees, but not all, are required to incorporate evaluation</td>
<td>13.7</td>
<td>50.0</td>
</tr>
<tr>
<td>The foundation conducts its own evaluations of grantees and funded programs</td>
<td>14.5</td>
<td>12.5</td>
</tr>
<tr>
<td>The foundation uses an outside consultant to evaluate grantees and funded programs</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

A shown above, CFs from the 90th percentile by endowment size (henceforth “large CFs”) were less likely to indicate that they had no formal requirement for evaluation (Table 13). However, it should also be noted that the large CFs also responded with greater frequency that there is no formal requirement for evaluation of grantee projects. The most popular response for large CFs was that some types of grantees, but not all, are required to incorporate evaluation.
Although the test for independence shown in Table 12 found a statistically
significant difference in the distribution of responses on question two of the survey, the
pattern of responses does not clearly support the theoretical expectations. The share of
CFs that have no formal evaluation requirements is only modestly smaller for large CFs.
Instead, it appears that large CFs are only more likely to have requirements for grantee
evaluation in some instances, which may simply reflect the fact that larger CFs have
more complex requirements and engage in a wider diversity of grant-making. However,
the details of when or why CFs selectively require evaluation for some grantees falls
beyond the scope of these survey responses.

The next survey item found to have differing responses by endowment size group
was question 5, part 1. The fifth question asked respondents to rate the degree to which
four different factors were present in the grantee evaluations conducted for the CF. The
general purpose of question 5 of the survey was to gauge the level of formality and use of
standards by CFs; therefore, the four factors listed in question 5 of the survey were
designed to correspond to the main categories of the Joint Committee (1994) evaluation
standards. It was expected that larger CFs might adhere to higher or more formal
standards for evaluation, compared to other CFs; however, only the first part of question
5, which refers to utility, was found to show a statistically significant difference in
response.

As shown in Table 14, more large CFs indicated that their evaluations either
"sometimes" or "almost always" involved the utility of the evaluation. None of the large
CFs responded either "never" or "rarely" and the portion that indicated that they "don't
know" whether utility was a factor in the evaluations they require was similar across CFs
of all sizes. This finding seems to suggest that large CFs are more interested in utility
standards in their evaluations.

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>17.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Rarely</td>
<td>18.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>20.5</td>
<td>37.5</td>
</tr>
<tr>
<td>Almost always</td>
<td>25.6</td>
<td>50.0</td>
</tr>
<tr>
<td>Don't know</td>
<td>17.9</td>
<td>12.5</td>
</tr>
</tbody>
</table>

The second part of question 5 also exhibited a statistically significant difference in
responses for large CFs compared to other CFs. The second part of question 5 addressed
another measure of formal evaluation practice standards by asking the surveyed CFs to
rate the presence of feasibility standards in grantee evaluations. As shown in Table 15,
the pattern for part 2 of question 5 was similar to what was seen in the responses to part
1. The majority of large CFs indicated that they almost always consider the presence of
feasibility when gauging grantee evaluations, compared to only approximately one-fourth
of all other CFs. At the other end of the scale, no large CFs gave a response of never or
rarely to the question.

On the overall issue of the inclusion of factors related to the major criteria areas
put forth by the Joint Committee (1994), however, the data is inconclusive. Of the four
parts of question 5 that address the standards, only two, the measures related to utility and
feasibility, were able to meet the standards for a statistically significant difference in responses between large CFs and other CFs, as was shown previously in Table 12.

Table 15

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>21.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Rarely</td>
<td>12.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Sometimes</td>
<td>19.7</td>
<td>25.0</td>
</tr>
<tr>
<td>Almost always</td>
<td>25.6</td>
<td>62.5</td>
</tr>
<tr>
<td>Don't know</td>
<td>20.5</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Although the differences for the two parts do fit theoretical expectations, which is to say that the factors were rated as being of greater importance by the large CFs, it is not clear that large CFs are more interested in standards overall. Instead, it is simply possible that large CFs are more concerned with specific types of criteria, specifically utility and feasibility, but they are not necessarily more concerned with evaluation standards overall.

The next item to demonstrate a statistically significant difference in responses between large CFs and other CFs was question 6, part 2. Question 6 of the survey asked respondents to rate the importance of eight activities to their CF’s mission. Several of the parts of question 6 addressed evaluation-related activities; however, part 2 asked about the importance of cultivating new donors. The results of large CFs and other CFs are shown in Table 16.
Table 16

Question 6 Part 2: Percent of Responses by Endowment Size Category

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not important</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Somewhat important</td>
<td>4.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Very important</td>
<td>13.1</td>
<td>25.0</td>
</tr>
<tr>
<td>Essential</td>
<td>82.8</td>
<td>62.5</td>
</tr>
<tr>
<td>N/A</td>
<td>0.0</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Although there is a statistically significant difference between the responses of the large CFs compared to other CFs, the practical significance is less clear. A smaller portion of the responding large CFs indicated that cultivation of new donors is essential, while a slightly larger share indicated that new donor cultivation is very important. No CFs of any size rated new donor cultivation as being not important. Overall, these findings suggest that large CFs may be slightly less concerned with cultivating new donors, which makes sense given the relatively large endowments these organizations already possess. However, this part of question 6 does not measure evaluation usage or interest and therefore neither supports nor rejects any part of the hypothesis regarding CF size and evaluation.

Question 7 of the survey asked respondents to indicate whether or not their CF’s staff holds college degrees in evaluation. Although the share of CFs responding to this question in the affirmative was somewhat low overall, a greater portion of large CFs answered yes than did the other CFs (Table 17).
The finding that large CFs are more likely to employ staff with formal evaluation degrees fits with the theoretical expectations discussed earlier. One possible explanation for this is that well-funded CFs may be more able to afford to hire professional evaluation staff and may employ more staff in general, compared to smaller CFs. Additionally, if CFs are more concerned with evaluating large grants than small grants, it seems logical to expect that the CFs most capable of providing large-dollar-value grants will be more likely to employ staff with evaluation training.

Of course, despite the difference that was identified between CFs with large endowments and other CFs, it is important to remember that the majority of CFs indicated that they do not employ any staff with college degrees in evaluation or assessment. While large CFs show more interest in hiring professional staff with evaluation degrees, or possibly simply exhibit a greater financial ability to hire such staff, the overall level of practice demonstrated by all CFs remains quite low. The majority of large CFs, three-fourths, do not employ staff with evaluation degrees. This may indicate either a lack of interest that is only slightly less widespread relative to smaller CFs, or a limited supply of workers with formal evaluation degrees in the marketplace.
A similar pattern was found on question 8, which queried respondents about whether or not their CF had received any formal training (but not a degree) on how to conduct evaluation. The majority of CFs employ staff that has received at least some level of formal evaluation training, although the majority is much greater amongst the larger CFs (Table 18). Respondents at larger CFs were also more likely to report that they did not know whether CF staff had received any evaluation training, which may simply reflect the difficulty of knowing about all the staff in a larger organization. Like the previous question, the response pattern to question 8 supports the theoretical expectation that large CFs will be more likely to engage in, or in this case support, evaluation than other CFs.

Table 18

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>53.3</td>
<td>75.0</td>
</tr>
<tr>
<td>No</td>
<td>42.6</td>
<td>6.3</td>
</tr>
<tr>
<td>Don't know</td>
<td>4.1</td>
<td>18.8</td>
</tr>
</tbody>
</table>

The last question to show a statistically significant difference in responses between large CFs and other CFs was question 11, which asked whether the CF had ever expanded funding to a grantee because of strong evaluation results. As shown in Table 19, a slightly higher share of large CFs indicated “yes” than other CFs, and a much smaller share of large CFs than other CFs indicated “no.” A sizable difference was also seen in the “don’t know” responses, which large CFs indicated at a higher rate than other
CFs. Unfortunately, this seems to suggest that large CFs are simply less likely to know whether or not additional funds were granted based on a positive evaluation, since the difference in the share of “yes” responses was small in practical terms. It may simply be that respondents that represent large CFs were less likely to know all of the answers regarding activities that have gone on in the organization because of its sheer size.

Table 19

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>44.2</td>
<td>50.0</td>
</tr>
<tr>
<td>No</td>
<td>48.3</td>
<td>25.0</td>
</tr>
<tr>
<td>Don't know</td>
<td>7.5</td>
<td>25.0</td>
</tr>
</tbody>
</table>

Overall, the survey items that showed a statistically significant difference in responses between large CFs and other CFs when using the 90th percentile endowment definition provide only limited support for the general hypothesis that evaluation practice and use would be greater among larger CFs than other CFs. First, only seven survey items, previously discussed in this section, showed statistically significant differences based on a p-value<0.05. Out of those seven, five showed response patterns that met theoretical expectations, which is to say that the large CFs responded in a way that indicated higher ratings to measures of evaluation practice or use. One of the survey items, question 11, did not clearly demonstrate a pattern of greater evaluation usage for large CFs, and the remaining survey item, question 6 part 2, was not an indicator of
evaluation usage. Therefore, the result is only weak support for the notion that large CFs have a different view on evaluation when endowment size is used to define size.

However, another definition of size was also used: population of the CF’s service area. The next section details the response patterns exhibited when the 90th percentile service area population size definition was used. As was shown earlier in Table 12, a total of four survey items met the standard of statistical significance, based on a p-value < 0.05; results for each item are detailed in the next section.

**Survey Response Differences of CFs with Large Service Area Populations**

The second measure of CF size that was examined for a response difference was the population of the service area served by the CF. When the 90th percentile was used to define large CFs and the individual survey items were analyzed to determine if differences existed between the responses of large CFs and other CFs, six items showed a statistically significant difference based on a p-value < 0.05, as was previously shown in Table 12. This section provides a detailed analysis of the response provided to each item by the two size categories of CF and examines whether or not the actual pattern displayed supports the previously stated hypothesis about the relationship between CF size and evaluation usage.

Theoretically, the expectation for any differentiation in responses between the large CFs and the other CFs in terms of population service area is identical to the expectation expressed for the comparison of responses by endowment size. Larger CFs are expected to be more likely to express an interest in evaluation or to more readily engage in evaluation than other CFs. If true, possible causes could include greater resources, better staff knowledge, pressure from donors, or intensity of competition for
grant assistance; however, the purpose of this analysis is not to determine why, but merely to assess whether the theoretical expectations have any merit.

The first survey item to show a statistically significant difference was question 2, which asked respondents to select the statement that best described their CF’s policy on evaluation. As shown in Table 20, the large CFs provided a mixed reading on whether or not they evaluate more than other CFs. A slightly higher share of large CF respondents than other respondents indicated that there were no formal evaluation requirements in place, 27.8 percent versus 24.3 percent. Large CFs were also less likely than other CFs to indicate that grantees in general are required to have an evaluation component and none of the large CFs indicated that they conduct their own evaluations of grantees or use outside evaluation consultants to assess grantee activities.

Instead, the largest share of large CFs, 44.4 percent, indicated that some, but not all grantees are required to incorporate evaluation into their activities. This suggests that among large CFs that do require evaluation, there may simply be more complex guidelines in place regarding which grantees should be evaluated. It could also reflect a greater diversity in the number and types of grants awarded by larger CFs relative to other CFs.

Still, the overall pattern of responses on survey question 2 does not provide support to the hypothesis that larger CFs will show a greater level of evaluation usage. There is no apparent pattern of higher usage exhibited by the larger CF respondents overall, while the percent of CFs with no formal requirement is similar across the size categories. The fact that large CFs indicated that some types of grantees are required to be evaluated only suggests that the large CFs that do require evaluation may take a
different approach to determining when and how such evaluations should be done.

However, the responses are insufficient to provide evidence that the level of evaluation is higher overall.

Table 20

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is no formal requirement</td>
<td>24.3</td>
<td>27.8</td>
</tr>
<tr>
<td>Proposals are assessed based on inclusion of an evaluation component</td>
<td>12.2</td>
<td>11.1</td>
</tr>
<tr>
<td>Grantees are required to have an evaluation component</td>
<td>32.2</td>
<td>16.7</td>
</tr>
<tr>
<td>Some types of grantees, but not all, are required to incorporate evaluation</td>
<td>13.9</td>
<td>44.4</td>
</tr>
<tr>
<td>The foundation conducts its own evaluations of grantees and funded programs</td>
<td>16.5</td>
<td>0.0</td>
</tr>
<tr>
<td>The foundation uses an outside consultant to evaluate grantees and funded programs</td>
<td>0.9</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The next survey item to show a statistically significant response distribution was question 3, part 8. Question 3 of the survey asked respondents to rate on a four point scale the degree that the CF used evaluation for each of 13 possible purposes. The purpose that was addressed in part 8 of question 3 was whether evaluation was used to fulfill requirements necessary to receive funding from other foundations or government agencies. Table 21 illustrates the differences observed in the responses the large CFs compared to the remaining other CFs.
Table 21

Question 3 Part 8: Percent of Responses by Service Area Size

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>41.5</td>
<td>16.7</td>
</tr>
<tr>
<td>Minimal</td>
<td>18.6</td>
<td>50.0</td>
</tr>
<tr>
<td>Frequent or typical</td>
<td>22.0</td>
<td>11.1</td>
</tr>
<tr>
<td>Almost always</td>
<td>17.8</td>
<td>22.2</td>
</tr>
</tbody>
</table>

Overall, a greater proportion of large CFs than other CFs indicated at least some usage of evaluation to fulfill the requirements of outside foundation or government funding sources. Only 16.7 percent of large CFs responded “not at all” to the question, compared to 41.5 percent of other CFs. Still, the pattern of responses offered little evidence that large CFs use evaluation for this purpose at a high level. Although few large CFs indicated no use, the most popular response for larger CFs was “minimal,” which suggests that this is not a strong use of evaluation. Additionally, the portion of large CFs that rated the usage level at the top two levels of the rating scale—“frequent or typical” and “almost always”—was actually slightly lower overall than it was for other CFs.

In order to support the aforementioned hypothesis about the relationship between evaluation usage and CF size, it would have been expected that the responses to part 8 would have shown consistently higher ratings for the large CF group. Instead the response pattern observed for part 8 was not clearly stronger for large CFs. Once again, this item, although statistically significant, does not provide the necessary support for the existence of differences across CFs when using the population service area definition of size.
The next survey item with an identified statistically significant difference in responses across CF size groups was question 7, which asked respondents to indicate whether or not their CF employs any staff that possesses a college degree in evaluation. Having staff with an evaluation-specific degree is somewhat rare across all CFs; however, as shown in Table 22, it appears to be somewhat more common in the large CF group: 27.8 percent versus only 6.7 percent in the other CFs group.

Table 22

Question 7: Percent of Responses by Service Area Size Category

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>6.7</td>
<td>27.8</td>
</tr>
<tr>
<td>No</td>
<td>88.3</td>
<td>66.7</td>
</tr>
<tr>
<td>Don't know</td>
<td>5.0</td>
<td>5.6</td>
</tr>
</tbody>
</table>

The responses to question 7 that are shown in Table 22 meet theoretical expectations in that this measure of evaluation practice is greater amongst large CFs than other CFs. The pattern was also similar to what was seen when the 90th percentile endowment size definition was used to categorize large CFs. In both instances roughly one-in-four large CF respondents indicated “yes,” which was far higher than the rate for other CFs.

Question 8 of the survey offers a similar measure of evaluation usage and also exhibited a statistically significant difference in responses between large CFs and other CFs. On question 8, respondents were asked to indicate whether any CF staff has
attended formal evaluation training, such as college courses or training seminars. Table 23 shows the percent indicating each response for the large CFs and other CFs.

Table 23

<table>
<thead>
<tr>
<th>Question 8: Percent of Responses by Service Area Size Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Don't know</td>
</tr>
</tbody>
</table>

Despite the statistically significant difference in responses, the actual pattern of responses does not provide support for the hypothesis that large CFs are more engaged in evaluation than other CFs. The percent of large CFs that indicated “yes” on question 8 is actually slightly smaller than it is for the other CFs: 72.2 percent versus 85.3 percent, respectively. In addition to lacking support for the hypothesis, the pattern of responses when the 90th percentile service area population is used also differs from the pattern seen when endowment size was used to define large CFs. As was illustrated earlier in Table 18, CFs with endowments at or above the 90th percentile indicated that they are more likely to employ staff that has received formal evaluation training than other CFs.

The last two survey items that showed statistically significant differences in response were question 10 and question 11. These two questions asked respondents about whether they had taken specific funding actions as a result of evaluation findings. The results of question 10, which asked if grantee funding had ever been cut because of poor evaluation findings, are shown below in Table 24.
Table 24

Question 10: Percent of Responses by Service Area Size Category

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>43.3</td>
<td>72.2</td>
</tr>
<tr>
<td>No</td>
<td>46.7</td>
<td>16.7</td>
</tr>
<tr>
<td>Don't know</td>
<td>10.0</td>
<td>11.1</td>
</tr>
</tbody>
</table>

The majority of large CFs, 72.2 percent, indicated that they had cut funding to a grantee as a result of poor evaluation findings, which was a far larger share than indicated by other CFs. This finding fits theoretical expectations and suggests that larger CFs may be more likely to use evaluation findings to guide funding decisions than smaller CFs.

Question 11, which asked survey respondents whether the CF had ever expanded funding as a result of positive evaluation findings, offers another measure of evaluation usage. However, unlike the previous survey item, the response findings offer only modest support for the hypothesis that large CFs are more likely to use evaluation than other CFs. Although a slightly greater proportion of large CFs than other CFs indicated “yes” to question 11, as shown in Table 25, the finding was clouded by the large share of large CFs that responded that they don’t know whether or not their CF had ever expanded funding to a grantee as a result of evaluation findings.

The responses shown in Table 25 is similar to the pattern previously observed when the 90th percentile endowment size was used to categorize the CFs (Table 19). It appears that respondents from large CFs are less likely to know how to answer question 11, which may reflect a difficulty in keeping track of what actions are taken within a
large organization or a separation of the evaluation and grant decision-making functions that is more inherent in larger organizations.

Table 25

<table>
<thead>
<tr>
<th>Response</th>
<th>Other CFs</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>44.1</td>
<td>50.0</td>
</tr>
<tr>
<td>No</td>
<td>49.2</td>
<td>22.2</td>
</tr>
<tr>
<td>Don't know</td>
<td>6.8</td>
<td>27.8</td>
</tr>
</tbody>
</table>

Findings from the Analysis by Endowment and Service Area Population

Overall, the categorical analysis of responses by size turned up only spotty support for the existence of a relationship between the size of CFs and evaluation usage. After dividing the survey respondents into groups of large and small CFs, as defined by the 90th percentile of CFs by endowment size and by size of the service area population, the analysis found statistically significant differences in responses across categories on only a minority of individual survey items. The chi-square and Fisher’s Exact tests of independence indicated that, in most instances, no difference exists between the responses of CFs categorized as “large” and those of smaller CFs, regardless of the method used to classify size.

In instances where a statistically significant difference was identified, the pattern of responses did not always provide support for the hypothesis that larger CFs will show evidence of more evaluation use than other CFs. This finding was true regardless of the method used to classify the size of the CF: endowment size or service area population.
However, in general the endowment definition provided stronger results than the service area population size definition. Using the endowment size definition, seven survey items were found to have a statistically significant difference and when the patterns were examined, six of the seven items were at least generally supportive of the hypothetical expectation. Using the service area population definition, only six survey items showed a statistically significant difference in responses, and of those, only four fit the general hypothetical expectations.

Overall, the findings from the comparative analysis of CFs by measures of organization size provide, at best, only weak support for the hypothesis that large CFs have a greater tendency to make concrete decisions based on evaluations and to make investments in evaluation. There are several possible reasons that size does not appear to be as strongly associated with evaluation usage as expected. Possibilities include measurement error, statistical error (i.e. Type I error), or the simple fact that evaluation is driven by factors other than size. For example, the survey questions may not be valid measures of evaluation usage, the methods used to define CF size may be flawed, or the relatively small sample size may be contributing to Type I error.

Given the limitations of the study, any of the aforementioned factors could be preventing the analysis from identifying a stronger relationship between size and evaluation usage. However, it is also possible that size simply does not play a strong factor in determining evaluation usage or it perhaps plays less of a factor than it once did. The general hypothesis that CF size was a driver in evaluation usage was based on the findings of researchers such as Pauly (2005), Kramer (2005), and others that have suggested CFs were dismissive of evaluation because of a lack of time, staff, or funds.
with which to conduct evaluations—all issues potentially associated with the size and financial well-being of the organization. However, as will be discussed in the next section, at least some of the CF community suggests that factors not related to size, such as internal leadership, may instead play a decisive role in whether or not a CF embraces evaluation.
CHAPTER V

QUALITATIVE ANALYSIS

The mail survey described in the previous chapter was focused on collecting primarily quantitative data on a broad, representative sample of CFs from across the country. After the completion of the mail survey, a subset of respondents participated in follow-up sessions, which were focused on collecting in-depth information about evaluation usage, barriers to evaluation, and ways of improving evaluation practice for CFs. The format of the interviews was semi-structured and open ended in order to balance the needs of collecting a similar set of information, while still allowing interviewees the opportunity to provide insights or explore topics that might not otherwise have been anticipated. A copy of the four primary questions discussed with each participant is included in the appendix.

Methodology

To participate in the follow-up interviews, respondents to the mail survey had to “opt in” by providing their name and contact information (either a telephone number or an email address) as an optional response at the end of the survey form. Each individual who provided contact information was contacted by both email and phone if possible, although in four cases the respondent provided only an email address, which limited the contact options. Initial contacts were made during the week of January 31, 2011; second contacts were made the following week and the interviews were all completed between February 3 and February 11, 2011.
Out of the 139 completed mail surveys received, 27, or 19.4 percent, indicated that they would be willing to participate in a follow up interview session. However, the number of individuals that were actually available and willing to participate in the follow up interviews was smaller because of nonresponse and refusals. Five of the 27 respondents that opted in provided incorrect contact information, which resulted in either reaching a wrong phone number and/or having an email message returned as undeliverable. Additionally, one individual provided a functional phone number but was reported by the person that answered as no longer being employed with the organization. Out of the remaining 21 individuals, 14 did not reply to phone and email contact or opted out of the follow up, and seven completed an interview, for a response rate of 33.3 percent.

Each telephone interview was conducted by the primary researcher and lasted approximately 45 to 60 minutes. In order to help ensure that interviewees would feel comfortable sharing both positive and negative aspects of evaluation usage within their specific organization, respondents were granted assurances that they would not be directly identified through attributable quotes. However, participants were also informed that quotations would be used but that they would not be attributed to specific individuals.

Interview notes were taken by the primary researcher and examined for information that could address the primary research questions of this study through two main qualitative approaches. First, the data collected during the interview was examined to determine if the perspectives of the interviewees supported the findings from the mail survey, which is a qualitative research process known as triangulation. Second, the data
were analyzed for explanatory stories on the possible causes of CF engagement in, or failure to engage in, evaluation practice and usage that would not otherwise be captured through theoretical exploration or quantitative survey data analysis. The findings from the analysis process are presented in the next two sections.

**Triangulation Support for Previous Findings**

The respondents interviewed during the follow-up sessions supported some, but not all, of the findings that arose from the quantitative analysis of data from the mail survey. Overall, the interviewees demonstrated a strong interest in evaluation and indicated that there was little connection between their own organizations size or endowment and the CF’s practice and use of evaluation, which is in line with the results of the mail survey. Where the results of the interviews seemed to diverge was in respect to the utilization of evaluation, which was higher in the interviews than in the mail survey. The difference in utilization rates is not necessarily unexpected, since those who elected to opt-in for the follow up discussions may have been more likely to have experience with evaluation usage that they were interested in sharing. Conversely, those that responded to the mail survey were drawn from the general population of CFs and were less likely as a group to actually practice and utilize evaluation on a consistent basis.

All of the participants in the follow-up interviews indicated that they have a strong interest in evaluation, which is not surprising since their interest generally also drove their willingness to participate in further research on the subject. Having an interest in evaluation corresponds with the finding from the mail survey that CFs overall show a fairly high level of interest in evaluation.
When asked about what types of evaluations were conducted and how the CF used evaluation results, however, the answers of respondents began to diverge from the findings from the mail survey. Nearly all the follow up interview participants described their organizations as being regular and strong practitioners and users of evaluation. For example, one respondent indicated that her CF conducts at least four different major types of evaluation, including a five-year collaborative assessment of community needs, special evaluation of larger grant initiatives, a required final report from most regular grantees, and a contracted external evaluation of the CF organization as a whole. Conversely, only one respondent went so far as to claim that “evaluation is not used as much as it could or should be,” although this respondent was still able to list evaluation activities including community-wide needs assessment and satisfaction surveys of affiliate organizations in the community.

Overall, the small and self-selected group of follow-up respondents demonstrated a far higher level of utilization than the overall population of CFs. The finding was not surprising; however, the responses of the interviewees could not be used as further evidence of a lack of evaluation practice or evaluation usage amongst the general population of CFs. The respondents did, however, offer some insight into why their organizations practice evaluation and why some CFs may be hesitant to engage in evaluation to the same degree, which are discussed later in this chapter.

Evidence from the follow-up interviews also indicated that respondents did not believe that size, whether measured by endowment or service area, plays a major role in the degree to which a CF engages in evaluation practice or usage. In the earlier chapter on the results of the survey of CFs, an analysis showed that two measures of size,
endowment size and the population of the CF’s service area, do not have a bearing on most measures of evaluation usage. Although the original hypothesis was that larger CFs would show a greater tendency to practice evaluation or use evaluation results relative to smaller CFs, findings from both the survey and the follow up interviews suggest that this hypothesis should be rejected.

Most of the follow-up respondents described their own organization’s endowments as being moderately sized and none of the participants represented a CF known to be above-average in endowment size (based on data from the Foundation Directory Online (Foundation Center, 2010) and the self-reported data collected through the mail survey), such as found in larger or wealthier urban areas or at well-established CFs such as the Cleveland Foundation. Despite the moderate endowments claimed by the respondents’ CFs, most demonstrated relatively extensive use of evaluation, including both the operation of multiple evaluation initiatives and the use of evaluation for both planning purposes and for organizational decision making. In short, the CFs that were represented in the follow up interviews were described as having fairly high levels of both evaluation practice and evaluation usage despite not being representative of CFs from the larger end of the scale in terms of endowment size.

The geographic areas served by the respondents’ CFs were more diverse, and ranged in size from a city with fewer than 100,000 to an entire state with more than one-half million residents. Unlike the broad survey of CFs, which did not show strong evidence of actual evaluation practice or usage, the follow up participants nearly all described themselves as being evaluation practitioners, as well as users of evaluation results. For example, the respondents from the largest and the smallest service areas that
participated in the follow-up interviews both listed multiple current evaluation practices at their respective CFs and generally described strong levels of organizational interest.

The qualitative interviews generally provided additional evidence to support a rejection of the hypothesis that geographic service-area size has a relationship with a CF’s tendency to practice evaluation or use evaluation results. For one, the levels of evaluation practice and usage reported by respondents showed no real variation; all but one reported high levels of both evaluation practice and evaluation usage. If the hypothesis were true, it would have been expected that there would have been a notable difference between the smaller and larger service-area CFs. Additionally, if the hypothesis were true it would be expected that the CFs in the sample, which does show a bias toward CFs with high levels of evaluation practice and usage, would also be biased toward only organizations serving larger-than-average service-areas compared to the overall population of CFs. However, neither of the aforementioned situations was true.

Of course, the primary role of the qualitative follow up interviews is not to triangulate the findings of the mail survey, but to provide some explanatory evidence on why the CFs that use evaluation do so and what barriers or challenges they face. The basic finding of the triangulation analysis, which supports the basic findings from the quantitative analysis of data from the survey, merely provides additional confirmation and support. In the next section, the results of an analysis of the qualitative data from the follow up interviews is presented in the context of providing a deeper understanding of why CFs do or do not use evaluation and what might be done to further encourage the practice and use of evaluation by CFs.
Follow Up Participant Perspectives on Evaluation Practice and Use

Even a small group of interviewees can present wide range of ideas and views on a subject; therefore, a methodology is required to derive some meaning from the responses provided to each question discussed during the interview process. One approach is to develop thematic categories consisting of a general idea that was supported or described from multiple perspectives. After completing the interview sessions and reviewing the notes from the sessions, five major themes were identified as being common across the interviewees.

1. Internal factors drive evaluation practice.
2. Breadth of practice exceeds depth of practice.
3. Usage of results varies more than practice of evaluation.
4. The barriers are recognized but seen as being in the past.
5. Resources and support are seen as ample.

A discussion organized by theme follows.

Internal Factors Drive Evaluation Practice

Although not all of the follow up interview participants were in a top leadership role for their CF, such as an executive director or president position, nearly all of the interviewees indicated a belief that evaluation practice and usage was driven, at least initially, by internal staff leaders. Only two interviewees specifically indicated that the CF board played a role in driving either the use of evaluation or practice of evaluation for the organization and none of the respondents indicated that outside stakeholders, such as donors or community members, were behind an initial drive for more evaluation practice. One respondent suggested that their organization does not use evaluation as much as
desired, but also indicated that the evaluation that is conducted is used by the board to examine the satisfaction of partners in the community, as well as to gauge community needs and conditions.

During the discussions, it seemed that interviewees did not believe that a strong push for evaluation arose from outside the organization. Instead, the practice and use of evaluation was described as being related to an internal culture. As one interviewee described it, the process is about “changing our own behavior through evaluation” and not about being pushed into evaluation by outside stakeholders. In most of the CFs represented by the interviewees, evaluation practice and use was originally spearheaded by a leader and over time the concept of being oriented toward evaluation spread to other levels of the organization.

When asked about the interest level of the community and of donors, stakeholders suggested that evaluation was something that the CF needed to sell to the donors, as opposed to the other way around. One interviewee went so far as to complain about the lack of interest, stating that “evaluation doesn’t sway donors so much yet,” which left the interviewee’s CF in the position of having to defend their evaluation practice and results to donors that are not convinced of the worth of evaluation. However, other interviewees saw stakeholder interest as improving, which made it valuable for the organization to be taking a lead in the practice instead of having to catch up to the demands of donors and community members. According to another interviewee, “we’re just getting the sense of [interest], that donors try to assess impact instead of just doing good.” In the situations described during the interviews, stakeholders were not driving the use and practice of
evaluation by CFs, but were instead catching up to the benefits of evaluation activities that were already being developed by these organizations.

**Breadth of Evaluation Practice Exceeds Depth**

The participants in the follow up interviews detailed a broad range of evaluation practices at their CFs. One interviewee started the conversation stating that “we evaluate in many different ways” and proceeded to list evaluation activities that included an assessment of donor services, a survey of local nonprofits, a five-year community needs assessment, a required evaluation report from funded programs, and an investment committee that evaluates the CF’s endowment performance. Another interviewee indicated that the CF employs “a whole variety” of different types of evaluation that cover the majority of funded projects, all special or larger initiatives, and community need, or basically everything except that they “don’t do formal external evaluation of the CF as a whole.” Even the respondent who indicated that their CF does not do as much evaluation as desired was able to list two different evaluation initiatives: a survey of affiliates that the CF works with on grants and an assessment of community indicators that the board uses to identify needs and organizational direction.

As the previous examples illustrate, the interview participants listed a broad variety of ongoing CF evaluation practices. Evaluation practices mentioned by the group ranged from the outcomes of funded projects to the assessment of specific foundation functions, such as donor relations or fund management. Additionally, the listed evaluation practices seemed to represent a wide range of purposes, concerns, and possible uses. For example, community assessments and indicator reports could be used to assess community needs and assist in the planning of future initiatives, while endowment
performance evaluations and satisfaction surveys could be useful in assessing the internal performance of the CF and its staff. Other evaluation practices, such as the evaluation of grantee outcomes, can potentially be used for multiple purposes, including assessing the CF's own grant making process, as well as for planning future funding allotments based on the past performance of grantee organizations.

The commonality of broad practices suggests that evaluation may be part of a broad approach that is either adapted or not adapted by CFs. Instead of focusing on one area of evaluation practice, such as the evaluation of grantee programs, the respondents seemed to indicate that information collection and evaluation are a part of their organizational culture. One respondent described the interest in evaluation as being part of wanting "to justify all we do" as a CF. Another respondent supported the overall importance of being evaluative by rhetorically asking "internally, how do you know you are being effective?" In both respondent examples, the notion expressed was that the purpose evaluation, for them, was somewhat self-evident; in short, being evaluative was seen by these interviewees as the only way to address the basic question of whether or not the CF was doing good.

**Barriers to Evaluation**

Of course, even though the CF representatives that participated in the follow up interviews demonstrated mostly strong levels of evaluation practice, the findings from the mail survey and from existing research both indicate that this is not universally the case. One reason for the disparity in evaluation practice and usage between different CFs may be that there are barriers that some organizations do not overcome. Even the interviewees, who mostly see themselves as using evaluation at a fairly high level,
acknowledged that there can be barriers to evaluation practice and usage. Some of the interviewees indicated that their own CFs had to overcome barriers to be successful, while other barriers were recognized as being potentially problematic even if they were not necessarily a problem for the interviewee’s own organization.

When queried about barriers most of the interviewees acknowledged that resource limitations or costs could be a barrier to evaluation. This was not a surprising finding given that the issue of the cost of evaluation and the limited ability of foundations to support the staff time necessary to conduct proper evaluations has been frequently cited as an issue in the research literature (see for example Ostrower, 2004; Pauly, 2005; Slater, Constantine, & Braverman, 2004). However, respondents from the follow up interviews did not widely indicate perceiving resource limitations as posing a significant barrier to their own CFs. One interviewee described the situation as follows.

I don’t know why [CFs] wouldn’t move from interest [in evaluation] to doing. A guess is it isn’t easy. It can be expensive. Ours is the third largest [in the state] so we’re not really big.

This interviewee acknowledged that the expense of evaluation was a recognizable barrier, but at the same time seemed to imply that it should not be an insurmountable barrier and indicated that their own CF was not large in terms of overall resources.

Another interview participant also acknowledged that evaluation is costly and indicated that the CF didn’t evaluate all grantee outcomes but instead focused on the larger funded programs. In this example, the interviewee described a barrier only to outcome evaluations, not other forms of internal and self-evaluation efforts that can be conducted using existing fixed resources. In the case of this CF, much of the data
collection and evaluation work, aside from evaluating grantee program outcomes, was viewed as a shared responsibility of staff.

Of course, utilizing staff is not a clear cut way around the cost of evaluation. For some CFs, a lack of staff resources is the more difficult barrier to overcome. For example, one interviewee revealed that his CF does require a report from grantees upon project completion but at the same time complained that there were not enough staff resources available to systematically review the reports. In this case, the barrier was not necessarily a lack of evaluative intent or interest, but a barrier between collecting data and doing the work necessary to make sense of the reports that were received and draw evaluative conclusions about grantee activities.

Another issue was that even when staff are available there can be a lack of interest in evaluation or a lack of training or capabilities to appropriately review evaluation data that is collected. An interviewee indicated that staffing levels were not a major problem, but that staff time and resistance were potential barriers, stating that “initially there can be shock, resistance” by staff during introduction of evaluation responsibilities. Some staff, especially if they were originally hired to perform jobs other than evaluate grantees, may simply struggle to deal with data or may feel that time spent on analysis takes away from the core charitable mission. Additionally, staff may also resist participating in evaluation duties if the results conflict with their own interests or pet projects. As another interviewee put it “some people just don’t take negatives,” in reference to the problem of CF staff pushing back or withdrawing from evaluation when the results are unfavorable.
However, overall, the group of follow up interview participants did not seem to view internal CF sources, such as staff reluctance or costs, as major barriers to evaluation. Even in cases where there was trepidation, the interviewees suggested that the issue was overcome and ultimately ceased to be a barrier to the organization. Of course, it must be remembered that these interviewees represent a group that was positively-biased toward evaluation and that reported having had success in implementing evaluation within their own organizations. Instead, external issues seemed to be more highly recognized as a source of barriers. As one interviewee describes it

There was no resistance internally [to evaluation]. The community is kind of like turning a ship to get them to understand. Getting nonprofit organizations in a position to do it is tough. It’s a lot easier for them to just count noses. For example, an art festival moving from just counting attendance to looking at economic impact is tough. We try not to penalize and push learning instead.

External barriers listed by interviewees included donor disinterest, grantee resistance, and a lack of understanding or support in the community.

Grantees were a major point of discussion among the interview group. According to more than one interviewee, the grantees that do the work that CFs fund can be more emotionally invested, which could be a source of resistance to evaluation. One interviewee offered the example of a grantee that had been operating a project for 20 years, which was found by the CF’s evaluation process not to be cost effective.

According to the interviewee “[the director of the program] was mad. But she did go raise money and turned the program around. Some others would have quit, but we got lucky, which is rare.” The example illustrates both the positive and negative aspects of
evaluation for grantees. In this case, a negative evaluation finding improved the program, but it was also acknowledge that this may not always be the case.

A source of resistance may stem from the fact that some grantees fear a negative evaluation that could result in their hard work being deemed ineffective. Grantees that receive a negative evaluation from a CF may react poorly, resist being evaluated, and possibly quit. But the positive is that an ineffective program could be identified through evaluation, which would have saved the CFs money for a more effective program. Alternately, negative evaluation findings, such as in the aforementioned example, could spur a major improvement in the program that will allow it to become more effective and continue operations.

Emotion can also have a strong impact on donors, who may react differently to CF evaluation depending on how it fits with their existing interests and beliefs. For some CFs, evaluation can be a positive tool for working with donors. For one interviewee, evaluation was seen as an effective tool for answering questions, since donors increasingly approach their CF for guidance. For another interviewee, evaluation results were something to be shared as part of a shift toward making longer-term connections with donors. In these instances, evaluation appeared to be a tool for conversation and for assuring external parties that the CF is interested in doing the best job possible with the funds entrusted to their care.

However, in some cases, emotion can have a stronger sway over donors than evaluation findings, which may not always correspond with existing beliefs. According to one interviewee “some [donors] are ruled by emotion, they don’t like facts” and prefer to direct their funds toward their preferred causes regardless of whether there is evidence
of effective programs or a strong community need. As another interviewee put it, “donors seem to be interested in impact but they don’t invest in it. There has maybe been a backlash [against evaluation]. They anecdotally believe in things.” In such cases, donors can undercut support for evaluation practice and evaluation-guided decision-making by restricting their donations and ignoring CF guidance.

Although Internal Revenue Service rules for CFs do not allow donors the level of control that would be necessary to directly block or dissuade a CF from practicing evaluation, a lack of donor support for the results that arise from evaluation can create a dilemma (Johnson & Jones, 1994). For example, several interviewees noted the rise in restricted and donor-advised funds, which are endowments within CFs that are set up to support only a specific type of activity or charity. These restricted and donor-advised funds can limit the ability of CFs to change the direction of grant making if, say, a community assessment shows a change in local needs or if an evaluation were to show that a program in a targeted area had not been effective. Conversely, unrestricted funds can be used in a broader fashion since they are not earmarked for any specific activity beyond the general mission of the CF.

**Resources for CF Evaluation**

The last question presented to interviewees during the follow up interviews was whether evaluation resources, such as checklists or guidelines tailored to the needs of CFs, were adequate or if additional support would be useful to promote greater evaluation practice and usage. Most interviewees offered few suggestions of evaluation resources that they felt were needed and indicated that they believed current resources to be adequate. The suggestion that additional assistance is not needed was not surprising,
given that the sample of interviewees was biased toward CFs that are already strong practitioners and users of evaluation, which suggests that the represented organizations have already been successful in finding or developing the tools and resources necessary for evaluation.

However, the interviewees did provide some evidence of areas where resource development might be useful, as well as some possible reasons why the development of CF-specific evaluation checklists or techniques might not be valuable to the field. An analysis of the interview responses found three main findings:

- A one-size-fits-all approach would be ineffective.
- Not having to start from scratch can be beneficial.
- Any effort would need to recognize a hierarchy of capacities that differ across CFs.

One possible enhancement to CF evaluation, which has been implemented successfully in other fields, is the development of a checklist or guideline that can offer specific steps for conducting evaluations. Stufflebeam (2000) suggests that checklists can be developed for most any specific field or task and will lead to higher quality evaluations. Examples of task and field specific checklists include Scriven’s (n.d) checklist for evaluating educational products, Persaud’s (2007) model for using cost analysis, and Patton’s (2003) checklist for qualitative evaluation methods; all of which addresses a specific field, methodology, or technique instead of acting as a general guideline, examples of which include Scriven’s (2007) Key Evaluation Checklist or Stufflebeam’s (2007) CIPP model checklist.
The follow up interview participants, however, did not provide support for the development of a checklist for CF evaluation. Although one interviewee acknowledged that some guidance could be helpful, since “the less you have to invent the better,” most were skeptical that a single checklist or guideline would be appropriate for the wide range of CFs and evaluation activities. Interviewees expressed concerns about whether CFs with smaller staff and financial capacities would be able to meet the standards of such a guideline or checklist, as well as the appropriateness of a fixed checklist to the evaluation of very different CF activities, such as scholarship awards or funding for an arts program that can have very different outcomes than a nonprofit program working on housing or community health.

Although interviewees were not supportive of checklists or other approaches that were viewed as being a “one-size-fits-all approach” there was some support voiced for a more general availability of evaluation resources. Instead of fixed tools or guidelines, interviewees expressed an interest in resources and tools that can be adapted or that can serve as learning tools for CFs to create their own evaluation processes. Respondents indicated that it was helpful to not have to “create everything from scratch” and listed sources and experts that had been helpful to their own organization. Examples of sources of evaluation support mentioned by interviewees included the Center for What Works, which specializes in creating benchmarking systems for nonprofit organizations; the Monitor Institute, which provides impact assessment and other consulting services; CF Insight, which also provides benchmarking and evaluation support; and the Council on Foundations, which is a membership organization for all types of foundations.
Interviewees also indicated that many of their CFs had contracted with consultants, ranging from the Rand Corporation to smaller independent organizations, in order to conduct evaluations of specific projects or to provide one-time evaluation services. These types of organizations were described as offering services customized to the specific needs of the CF and being helpful as the CFs develop their own evaluation practices.

Finally, a concern of the interviewees was that outside evaluation tools would not properly recognize the large differences in evaluation capacity across CFs. There was some concern that evaluation standards or tools developed by larger or more evaluation-centric CFs would not have an impact on smaller CFs or those still struggling to develop some basic evaluation capacity. One interviewee suggested that the development of a broad “theory of change” that could be used by CFs would make evaluation “much more tenable” as opposed to the development of something more specific or standards oriented, such as a checklist.

Summary of Follow Up Interview Findings

The follow up interviews conducted after the mail survey of CFs provided two major areas of evidence to the study of CF evaluation practice. First, the follow up interviews acted as an additional source of empirical support for the some of the findings identified from the analysis of the survey data. Second, the follow up interview conversations were a source of detailed data on the drivers of evaluation practice and usage at CFs, as well as the barriers that CFs in general may need to overcome to improve the level of overall evaluation practice and usage in the field.
The follow-up interviews suggest that evaluation practice and usage is not driven by CF traits that can be easily identified from the outside. Instead, CFs that adopted stronger levels of evaluation practice and usage described an internal culture of evaluation, which was typically championed, at least initially, by a staff leader or board member. According to the interviewees, long-lasting evaluation practice and usage was described as being supported by a noticeable return on the investment made by the CF in evaluation, which is to say results that improved the performance of grantees, impressed donors, or assisted in the internal functioning of the organization. In short, the interviewees mostly described the evaluation practices of their own CFs as being sensible and not particularly burdensome nor necessitated by any outside demand.

Finally, those CFs that have already integrated evaluation did not see the barriers to evaluation as particularly challenging. The interviewees, although admittedly biased toward support for evaluation, mostly described barriers toward evaluation, such as resistance from grantees and internal stakeholders or resource limitations, as being surmountable by the typical CF. A number of existing resources for technical support and evaluation consulting were cited and there was little demand expressed for the creation of new tools or support, such as checklists or guidelines specific to the CF field.
CHAPTER VI

CONCLUSIONS

This study set out to address two basic questions about evaluation practice and usage by community foundations. The first research question was whether or not CFs actually practice and use evaluation to a degree related to their expressed interest in and support of evaluation. The second research question was whether any traits were associated with CF use or non-use of evaluation, which will be important for CFs and policymakers to know if they wish to encourage greater evaluation usage. To address these main questions, this study has used multiple research techniques, including a review of existing research, quantitative research using a nationally representative sample of CFs, and in-depth qualitative research, to test multiple hypotheses underlying the primary research questions.

Connection of Literature and Research Findings

Existing research provides only a basic understanding the current state of CFs and the issues faced in practicing and using evaluation. As discussed in Chapter III, the literature has generally drawn positive conclusions on both the potential value of evaluation to CFs and the trend of CF interest in evaluation. However, the evidence regarding measures of actual evaluation practice and usage has been scarce, with many researchers suggesting that there is a misalignment between interest in evaluation and actual practice or usage of evaluation.
The main assertion put forth by Pauly (2005), Ostrower (2006), Seita (1993) and others was that CFs consider evaluation to be important but at the same time they do not use evaluation at nearly the expected level. This study tested this assertion by querying CFs regarding measures of evaluation practice. The findings from the analysis of the results of the mail survey of CFs conducted during this study confirm the finding that actual evaluation usage is low. Approximately one-quarter of respondents indicate that they did not require any evaluation for grantees (Table 4). Although a majority do some type of evaluation for grantees, most other forms of evaluation remain relatively uncommon. For example, fewer than half of CFs conduct any type of community needs assessments and only about one-third indicated that they evaluate their own internal activities (Table 2). This suggests that, as a group, CFs have not fully embraced evaluation.

While conducting evaluations is important, the benefit of evaluating will not be realized if the results are not used to direct the organization’s decision making. Although a majority of CFs either conducts or requires at least some level of grantee evaluation, the evidence suggests that usage of the results may be limited. As discussed earlier, nearly half of CFs indicated that evaluation decisions are not used to either make decisions on grantee funding or even to promote the activities of the CF to donors or the community (Table 10). The disparity between the share of CFs that evaluate and the share that use evaluation results for funding decisions or promotion suggests that at least some portion of CFs is missing out on the benefit of evaluations that are already being conducted.

Findings regarding the reasons that CFs do not evaluate more are complex. The review of existing literature suggested that evaluation costs, difficulty in measuring
program outcomes, and the perceived threat of evaluation results to the CF organization and staff all play a role in discouraging evaluation. Although the purpose of this study was not to refute these prior findings, the results do offer some evidence to suggest that the challenges found by prior research may be overstated. For example, the analysis conducted using data from the survey shows that CFs with smaller levels of endowed assets did not conduct evaluation at a level below larger CFs. In other words, financial conditions may not drive evaluation as much as previously thought or, conversely, limited financial resources may not be as much of a barrier as some have suggested.

Additionally, qualitative discussion sessions with self-selected CFs, which were mostly high users of evaluation, suggests that internal factors such as leadership can overcome the main barriers CFs often cite as limiting their evaluation usage. While some of the existing literature has discussed why staff may not be supportive of evaluation, this study revealed at least some CF staff have been the driving force in promoting evaluation within their own organizations. Furthermore, if evaluation results are viewed as threatening by some CFs, as has been suggested, the condition is clearly not universal; many survey participants report evaluating and the participants in the follow-up interviews spoke highly of the benefits of knowing more about the outcomes the organization and its grantees are able to provide in the community.

In summary, this study confirmed that a large portion of CFs currently still do not evaluate and use evaluation, despite the fact that researchers such as Ostrower (2005) have shown that foundation leaders indicate increasing levels of interest in evaluation and accountability. The findings of the study also suggest that the barriers to evaluation practice and use for CFs that are discussed by the literature can be overcome and that
evaluation can be successfully used by different types of CF. When considered jointly, these findings bring forth the concern that CFs appear to be giving vocal support to evaluation, but shying away from actual practice and use of evaluation for reasons that remain unknown. Although the CFs that participated in the study did confirm that the barriers to evaluation still exist, they also suggested that evaluation can be successfully brought into the CF environment by an executive leader that supports the practice.

**Theoretical Hypotheses and Study Findings**

Existing literature on CFs and evaluation also provided a theoretical basis for testing a hypothetical link between select characteristics of the CF organization and the tendency to practice and use evaluation. As discussed earlier, this study examined the relationship by looking for statistical differences on measures of evaluation practice and evaluation usage in the mail survey and whether the CF was large, based on asset size or service area population size. The hypothesis would predict that larger CFs should show a greater tendency to engage in evaluation than smaller CFs.

As was shown in the results of the mail survey analysis in Chapter IV, on most survey items there was not a statistically significant difference in the distribution of responses between the large CF category and other CFs. Taken as a whole, the few measures that showed differing response distributions were not sufficient, in practical terms, to support a statement that larger CFs are more evaluative than other CFs. Although it is possible that the limited degree of support identified in these tests was hampered by a positive response bias where more evaluative CFs responded, it is not clear that such a positive bias would alter the overall findings. If evaluative CFs of all sizes were more likely to respond, then the overall difference between sizes should still
theoretically persist. If the response was somehow greater from large CFs that evaluate, then the smaller CFs should also still, in theory, respond in a way that would demonstrate their lower rates of evaluation practice and usage. The overall results would only be invalidated if the bias were such that only smaller CFs that have a strong tendency to evaluate somehow responded at a higher rate, which appears to be a scenario that is possible but unlikely.

Although overall support for size-based differences is weak, there was some evidence to suggest that larger CFs are more likely to employ workers that have received some formal evaluation training. Of course, any tendency toward heavier investment in evaluation training for staff may simply reflect that larger CFs are more likely to invest in training of all types for staff members. It is also possible that large CFs are simply more likely to offer to pay for employee training and therefore a greater number of the staff of large CFs have received training in evaluation, as well as other subject areas.

The follow up discussions also did not provide strong support for any theoretical assumptions regarding a relationship between organization size and evaluation use or practice. Although the follow up interview group was not representative of the full population of CFs, it should be noted that interviewees did not mention organizational characteristics or size as playing a role in the decision to practice evaluation. Instead, interviewees focused mainly on internal characteristics of the CF, such as leadership or organizational culture, which they felt had a stronger impact on the decision of the CF to become involved in evaluation.

Overall, this study failed to offer support for the hypothesized relationship between organizational size and evaluation practice or usage, as measured by the items
developed for the mail survey. This finding does not eliminate the possibility that there may be a relationship between some other organizational characteristic and evaluation; however, the measures used in this study did not lend strong support to the role of asset or service area size in driving evaluation practice or use.

The lack of a connection between CF size and CF evaluation usage also lends indirect support to the assertions of the small follow-up group, wherein interviewees suggested that barriers to evaluation usage could be overcome by a typical CF with the right leadership. Because size does not appear to play a large role in CF evaluation practice and usage, it suggests that the presence or lack of resources is not a major determinant of the ability to evaluate. In other words, it did not seem to be the case that just the big, well-financed CFs evaluated while the smaller, poorer CFs did not. This suggests that evaluation usage need not be limited by resources and supports earlier assertions that evaluation should be able to provide benefits to a wide range of organizations.

**Contribution to Literature**

This study has addressed several issues relevant to CFs and to the evaluation field that have not been addressed previously. First, this study helps fill in a knowledge gap regarding evaluation practice by CFs. Previous research in this area has typically been either very limited in scope or very broad in scope. For example, Seita’s (1993) study took a similar approach by using survey methodology to assess evaluation usage; however, the sample was geographically limited to Michigan while being open to types of foundations other than CFs. Similarly, Ostrower’s (2005) extensive survey of foundations was national in scope but included all types of foundations and did not focus
solely on evaluation usage. This study examined a national sample of CFs; unlike other studies its results can be more appropriately generalized to the full realm of CFs nationwide.

Other research, such as that conducted by Pauly (2005), Ostrower (2007), and Carson (2000) focused on developing a problem statement and case for evaluation usage for foundations, as opposed to providing a simple empirically-based assessment of the current state of evaluation within CFs. This study has developed data on the current status of evaluation in CFs, while also examining the research findings in the light of existing theory through the testing of basic hypothesis about organizational size and evaluation. In short, despite concerns of a possible positive bias, the estimates provided in Chapter IV provide some of the only current measures of evaluation practice and usage that are generalizable to all CFs in the nation. If nothing else, these estimates of non-practice and non-use could be considered a conservative low-end of the range of the true level occurring across all CFs.

Finally, through qualitative follow up interviews this study also offers a cursory exploration of the reasons why CFs use evaluation and how evaluation can be integrated into the CF environment. Although the follow up sessions were small in both participant size and scope, the interviewees brought to light the importance of staff leadership in bringing evaluation into the CF environment. This finding provides a new avenue to consider that stands in contrast to some of the existing research that has suggested factors such as cost or evaluation capacity were standing as the primary barrier to increased levels of evaluation practice in the field of philanthropy.
Lessons for Community Foundations

The focus of this study has been on what CFs actually do in terms of evaluation and the degree to which evaluation is associated with organizational size and assets; however, the findings also support discussion of some general statements about what CFs can, or possibly should, do in terms of practicing evaluation. First and foremost, CFs need to align actual usage of evaluation levels with support for the practice. In short, this study found levels of evaluation practice that do not match the relatively high levels of interest in evaluation that other researchers, such as Ostrower (2005), have previously documented. If evaluation is important, as most say, then it should move beyond being a worthy goal to being a regular, integral part of the philanthropic process.

CFs that are considering whether evaluation is worthwhile should consider the potential uses, and therefore benefits, of evaluation. For example, Table 5 lists a dozen uses of evaluation that at least some CFs reported engaging in at some level. Logically, it would seem that information from evaluations could be useful for deciding on grantees that should receive more funding; however, the majority of respondents indicated that this is done only “minimally” or “not at all.” Although more research would be helpful in order to provide estimates of the additional benefits that could be derived from extending evaluation practice and use, the expansion of evaluation for additional internal decision making purposes represents one way that CFs might quickly expand their own benefits from the practice without necessarily incurring added costs.

Additionally, CFs should consider taking a broader mindset when considering the parties that might benefit from evaluation results. Approximately one-in-five CFs do not use evaluations to offer help to local non-profit agencies, which are sizeable category of
grantee for many CFs (Table 5). Although CFs may primarily think of evaluation as something to be used internally by the board or staff, there would seem to be little downside to sharing evaluation findings that could be useful to relevant grantee organizations. Non-profit agencies that are evaluated as part of their grant activities with a CF could benefit from detailed evaluative feedback and could potentially help these organizations find more effective ways of working in the community, even if it is outside the activity funded by the CF. Furthermore, evaluations can contribute evidence on what works and does not work in terms of programmatic approaches to addressing community issues.

Similarly, CFs do not always use evaluation to promote their own work in the community or to donors. Sharing evaluation results could potentially be a way of demonstrating accountability and trustworthiness to donors and a commitment to doing the best possible work for the community. The survey found that only about half of CFs currently use evaluation to promote their programs to the community, and fewer than half use evaluation to demonstrate outcomes to donors (Table 5). The relatively low showing suggests there are still many CFs that could potentially gain some benefit from using evaluations to help communicate their efforts in a more public way. Indeed, showing a commitment to evaluation within the CF could potentially be a first step to building external support for evaluation from donors and the community at large. If a CF can show that evaluation extends a benefit, such as better outcomes or more efficient programs, donors may be more interested in directing funds toward both the practice of evaluation and the initiatives that receive high marks from subsequent evaluations.
Finally, for those CF staff or board members who have the desire to transform their own foundation into a more evaluative organization, the suggestions of respondents who participated in the interview sessions should offer encouragement. Among CFs that consider themselves to be active practitioners and users of evaluation, internal leadership was often cited as being a driving force in changing the way the organization operated in terms of evaluation. Individual leaders within CFs can be one path to bringing a more rigorous evaluation approach into acceptance.

**Suggestions for Future Research**

This study provides only a basic look at the current state of evaluation in CFs and the possible factors that contribute to greater practice and use of evaluation. Additional research is needed to understand the causes of uptake of evaluation practice by CFs and the benefits of evaluation to CFs. One of the arguments for increasing evaluation in the CF environment has been that evaluation should constitute a way that CFs can add value to donors and communities beyond the role of investing and maintaining funds (Porter & Kramer, 1999). Additional support for or against this argument is needed in the form of estimates of the benefit of evaluation, such as cost savings associated with choosing better programs and grantees, or increased beneficial outcomes associated with choosing the most pressing community problems or the most effective program to address an issue. Convincing CFs to invest in evaluation may first require developing empirical evidence that supports a tangible benefit of evaluation to one or more groups of stakeholders, since the current status of CFs may make the cost or risk of evaluation much more apparent than an unknown reward.
Another venue for future research is the further investigation into the relationship between organizational characteristics and the level of evaluation practice and use by CFs. This study examined only the relationship between CF size, as measured by two definitions, and multiple measures of evaluation practice and usage; however, there are many additional organizational characteristics that could potentially play a role in whether or not evaluation is adopted by a CF. Examples of factors that could warrant consideration for future research include staff size, board composition, geographic region, rural versus urban service area, age of the CF, and community demographics. If one or more factors could be identified as having a relationship with the implementation of evaluation, it could help identify barriers that CFs face to evaluation and provide insight into how greater evaluation practice could be encouraged.

Finally, further research should also be conducted into the specific evaluation needs of CFs and whether they differ from other types of foundations and nonprofit organizations. Although this study did touch on CF evaluation needs through the qualitative follow-up interviews, there were few suggestions for tools or other forms of assistance that might be needed to help CFs more easily conduct evaluations. Part of the reason for the lack of suggestions may have been the nature of the small sample of interviewees, who primarily represented organizations that were already engaged in evaluation. Ultimately, pushing evaluation practice and use forward in the CF realm will require both determining what an ideal level of evaluation might be and identifying the steps necessary to reach that level. Much can still be learned from other fields such as education, government, and even business, where evaluation is more widely used.
Lessons Learned

Finally, in addition to the previously discussed findings on CF evaluation practice and usage, the results of this study also offer lessons about the process and challenge of conducting research on CFs and evaluation, as well as in a broader sense for research on organizations. Over the course of this study, several issues were encountered that could potentially impact similar future research, and that could potentially be addressed through adjustments in method or design. These issues of concern include uncertainty about the definition of evaluation understood by CFs and the relatively small response rates, both for the mail survey and amongst those who initially opted-in to participate in the follow-up survey.

Evaluation is typically defined within the field as being the systematic determination of merit, worth, or significance (Scriven, 1991). However, because evaluation practice may not be as well defined to CFs, it is possible that respondents held differing definitions of evaluation, which could potentially affect the nature of responses on items intended as measures of evaluation practice or usage. This issue may be partially addressed by examining the reliability of responses across respondents; if reliability is high, it would suggest that those respondents held a common understanding of evaluation. However, respondents who do not hold the same definition of evaluation may have responded to the questions in a very different way or not at all. To address this, future research on evaluation should offer both a definition and examples of what evaluation is when questioning individuals or organizations that may not be familiar with an established definition.
A larger issue for this study, and perhaps for survey-based research in general, is response rate. As discussed earlier, only 15.7 percent of surveyed CFs responded to the mail survey and only about one-third of those that initially expressed interest in completing a follow-up interview session ultimately participated. Although response rates are a well-known challenge for researchers, there are several techniques which could potentially have improved the response rate for the survey used in this study.

One technique that can boost response rates is multiple contacts (Czaja & Blair, 2005). This study employed only one survey mailing; however, additional follow-up contact would have most likely encouraged additional CFs to respond. A second or third mailing to non-respondents is a typical approach, however, if email addresses or phone numbers were available, it would have potentially also been possible to offer CFs other ways of participating. Although costs rise with multiple contacts, the use of random selection can help by selecting a small sample of the survey population that is then intently focused on in order to obtain the desired response. The number of surveys received would not necessarily vary from what was obtained in this survey, but the response rate from the small sample would be higher, while a randomization process would help ensure that the sample was representative.

Past surveys of CF have also shown higher response rates by obtaining a letter of introduction from an entity that is known and trusted by those being surveyed (for example Seita, 1993). Ideally, an organization such as a well-known CF or a trade or industry group could have potentially offered credibility to the survey. Although this study was not able to obtain such an endorsement, future surveys should consider finding organizations that could help encourage respondents to participate, especially considering
that an endorsement letter could come at essentially no cost if the research has an established relationship with the endorsing organization.

Finally, both the issues of timing and the response preferences of those being surveyed are also possible contributors to response rate that warrant consideration in future CF and organizational survey work. The survey in this study was conducted only by mail; however, there may have been some benefit to offering additional ways for participation, such the option to respond via a web site or phone call. Each method of contact—mail, email/online, and phone—holds advantages and disadvantages; however, response may potentially be maximized by using the maximum amount of contact techniques that are reasonable and affordable in the situation.

Timing of the survey should also be considered. The mail survey was conducted during October and November, which left the follow-up contacts to occur during the holiday season. The initial follow-up contacts did not obtain a strong response and further phone and email contact was necessary before successful contacts were made and interview sessions were ultimately conducted as long as two-months later. Avoiding the holiday period and promptly conducting follow-up contacts and interview sessions might have also helped improve the completion rate for the interview sessions.
REFERENCES


Appendix A

Survey of Community Foundations and Evaluation Usage
Survey of Community Foundations and Evaluation Usage

Thank you for taking a few minutes to complete this brief survey on the use of evaluation by community foundations. The results will help provide information on how community foundations practice evaluation, as well as how evaluation helps foundations bring value to the communities they serve. All individual responses will be kept confidential and no identifying information on individual organizations will be revealed in the final report.

Please return your completed survey in the enclosed postage-paid envelope by Friday August 27, 2010.

This research is being conducted as part of a Doctoral program through The Evaluation Center at Western Michigan University. If you have any questions, please contact Brad Watts by email at brad.r.watts@wmich.edu or by phone at 269-216-4707.

1. Does your foundation have in place systematic procedures to conduct any of the following types of evaluations? Please check all that apply.

   o Assessment of the needs of the community you serve (e.g. research into local social or economic issues)
   o Evaluation of applications from individuals seeking grant funding or financial support
   o Evaluation of applications for funding from organizations
   o Evaluation of the activities of non-profit organizations funded by your foundation
   o Evaluation of the outcomes of projects funded by your foundation
   o Evaluation of the effectiveness of your foundation’s own activities
   o Other (please describe)

2. To what degree, if any, does your foundation encourage or require grantees to practice evaluation? Please check ONLY the one answer that best applies.

   o There is no formal requirement; however, our guidelines encourage grantees to conduct their own evaluations
   o Proposals or requests for funding are assessed based on whether or not they include an evaluation component
   o All grantees are required to have an evaluation component
   o Some types of grantees, but not all, are required to incorporate evaluation to receive funding
   o The foundation conducts its own evaluation of grantees and funded programs
   o The foundation uses an outside consultant to evaluate grantees and funded programs
3. On a scale of 1 to 4 please rate the degree to which your foundation uses evaluation or the results of evaluations for the following purposes. (Rating scale: 1 = not used at all, 2 = used only minimally or occasionally for this purpose, 3 = a frequent or typical use of evaluation, 4 = essential; evaluation is always used for this purpose)

<table>
<thead>
<tr>
<th>Rating of Usage</th>
<th>Purpose of Evaluation Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td></td>
</tr>
<tr>
<td>Minimal</td>
<td></td>
</tr>
<tr>
<td>Frequent or typical</td>
<td></td>
</tr>
<tr>
<td>Almost always</td>
<td>To approve or select grant applicants for financial support</td>
</tr>
<tr>
<td></td>
<td>To identify the most pressing problems in the community</td>
</tr>
<tr>
<td></td>
<td>To assess our own operations</td>
</tr>
<tr>
<td></td>
<td>To support fundraising efforts or to show results to donors</td>
</tr>
<tr>
<td></td>
<td>To provide board members with information for decision making</td>
</tr>
<tr>
<td></td>
<td>To help local non-profit organizations improve their operations</td>
</tr>
<tr>
<td></td>
<td>To publicize our foundation's achievements to the community</td>
</tr>
<tr>
<td></td>
<td>To fulfill the requirements to receive funding from other foundations or government agencies</td>
</tr>
<tr>
<td></td>
<td>To ensure that organizations/programs funded by the foundation meet legal requirements</td>
</tr>
<tr>
<td></td>
<td>To generate information on the validity of a type of program or theory (e.g., to demonstrate the effectiveness of a type of poverty alleviation program)</td>
</tr>
<tr>
<td></td>
<td>To select programs or funding opportunities that should receive additional funding</td>
</tr>
<tr>
<td></td>
<td>Other (describe)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
4. Does your foundation use or promote the use of specific standards or guidelines for evaluation? (Please select only the one answer that best fits)

○ Grantees are not required to be evaluated and therefore no guidelines are provided
○ Grantees and programs are required to be evaluated, but we do not specify any specific standards or guidelines for how that evaluation should be conducted
○ Sometimes specific guidelines or checklists are recommended or required, but it depends on the situation
○ We recommend or require the use of a standard set of guidelines or checklist that has been created by an outside organization (e.g. Key Evaluation Checklist, Joint Committee on Standards, the CIPP model, etc.)
○ We recommend or require the use of evaluation guidelines or standards that were specially developed by our foundation
○ Other:

5. Please rate the degree to which each of the following factors in conducting an evaluation of grantees is present in the evaluations conducted for your foundation. (1 = never present or important, 2 = rarely present or important, 3 = sometimes present or important, 4 = almost always present or important, DK = Don’t know)

<table>
<thead>
<tr>
<th>Factors considered when evaluating grantees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility of the evaluation, meaning that all key stakeholders have been identified and those conducting the evaluation are credible</td>
</tr>
<tr>
<td>Feasibility, meaning that the evaluation is non-disruptive, sensitive to political issues, and cost-effective</td>
</tr>
<tr>
<td>Meets basic standards meaning that individual rights are respected, strengths and weaknesses are fairly reported, findings are fully disclosed, and conflicts of interest are avoided</td>
</tr>
<tr>
<td>Accuracy is monitored, meaning that grantees’ activities are documented, information is collected in a systematic manner, appropriate analysis techniques are utilized, conclusions are justified by data, and reporting is performed in an impartial manner</td>
</tr>
</tbody>
</table>

6. On a scale of 1 to 4 (1 = not important, 2 = disagree, 3 = agree, 4 = strongly agree) please rate the importance of the following activities as they apply to your Community Foundation’s mission.
<table>
<thead>
<tr>
<th>Not Important</th>
<th>Somewhat important</th>
<th>Very important</th>
<th>Essential</th>
<th>Not Applicable</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Managing endowment funds</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Cultivating new donors</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Assessing community needs and identifying the most pressing problems</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Selecting grantees that will have the greatest impact on community issues</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Acting as a resource for community information</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Supporting the causes that are most important to major donors</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Ensuring that charitable funds are utilized by grantees in an efficient manner</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>n.a.</td>
<td>Determining the impact of programs funded by the foundation on local community</td>
</tr>
</tbody>
</table>

For Questions 7 – 13 please circle one response: yes, no, or don’t know

7. Do any members of the community foundation’s staff hold college degrees in evaluation or assessment? Yes No Don’t know

8. Have any members of the community foundation’s staff attended formal training, such as seminars or college courses, on how to conduct evaluation? Yes No Don’t know

9. Have you ever promoted your Community Foundation’s evaluation or community needs assessment activities to donors? Yes No Don’t know

10. Has your Community Foundation ever cut funding to a grantee because of poor evaluation results? (i.e. missing outcome goals, not showing improvement) Yes No Don’t know

11. Has your Community Foundation ever expanded funding for a grantee because of strong evaluation results? Yes No Don’t know

12. How big is your Community Foundation (estimated current endowment): $___________ million

13. How big is the community you serve (est current population of the area): _______________

14. [OPTIONAL] Would you be willing to participate in a follow up phone interview or be part of a case study? If so, please provide your name and preferred contact information:

Name ______________________ Phone __________ Email __________
15. Do you have any comments about evaluation in your community foundation or about community foundations in general that might help us to understand the usage and importance of evaluation activities?

Thank You for your response!
Appendix B

Survey Follow-Up
Thank you for agreeing to participate in a brief follow-up to the survey on use of evaluation that you participated in last Fall. The goal of the following set of questions is to gain a better understanding of how better evaluation practice and tools might possibly assist community foundations, as well as the organizations and individuals supported by community foundations, in achieving their goals.

This research is being conducted by Brad Watts as part of the Interdisciplinary PhD in Evaluation program at Western Michigan University. If you have any questions you can contact me at brad.r.watts@wmich.edu or by phone at 269-216-4707. Thank you!

Questions for Community Foundations

1. Can you describe how your organization currently uses evaluation? For example, explain at what steps in the granting and fundraising processes is evaluation used? For what purposes are evaluation results currently used?

2. The initial survey results suggested that most community foundations are interested in evaluation, however, many are not engaging in evaluation at the desired level. Considering your own organization, do you think that evaluation is used as much as it could be? Why or why not?

3. Are there any barriers to conducting evaluation or using evaluation results that you view as being unique to community foundations as a type of organization?

4. Would a checklist or systematic set of guidelines that was designed to help community foundations in conducting evaluation be helpful? If so, please describe what elements or features you would like to see.
Appendix C

HSIRB Approval Letter
Date: July 20, 2010

To: Brad R. Watts, Student Investigator

From: Amy Naugle, Ph.D., Chair

Re: Approval not needed

This letter will serve as confirmation that your project “Community Foundations and Evaluation Usage” has been reviewed by the Human Subjects Institutional Review Board (HSIRB). Based on that review, the HSIRB has determined that approval is not required for you to conduct this project because you are collecting information about foundations and are not collecting personal information about individuals. Thank you for your concerns about protecting the rights and welfare of human subjects.

A copy of your protocol and a copy of this letter will be maintained in the HSIRB files.
Appendix D

Census Bureau Regions and Divisions with State FIPS Codes
Census Bureau Regions and Divisions with State FIPS Codes

Region 1 Northeast

<table>
<thead>
<tr>
<th>Division 1</th>
<th>New England</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connecticut (09)</td>
<td></td>
</tr>
<tr>
<td>Maine (23)</td>
<td></td>
</tr>
<tr>
<td>Massachusetts (25)</td>
<td></td>
</tr>
<tr>
<td>New Hampshire (33)</td>
<td></td>
</tr>
<tr>
<td>Rhode Island (44)</td>
<td></td>
</tr>
<tr>
<td>Vermont (50)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Division 2</th>
<th>Middle Atlantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Jersey (34)</td>
<td></td>
</tr>
<tr>
<td>New York (36)</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania (42)</td>
<td></td>
</tr>
</tbody>
</table>

Region 2 Midwest

<table>
<thead>
<tr>
<th>Division 3</th>
<th>East North Central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indiana (18)</td>
<td></td>
</tr>
<tr>
<td>Illinois (17)</td>
<td></td>
</tr>
<tr>
<td>Michigan (26)</td>
<td></td>
</tr>
<tr>
<td>Ohio (22)</td>
<td></td>
</tr>
<tr>
<td>Wisconsin (55)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Division 4</th>
<th>West North Central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iowa (19)</td>
<td></td>
</tr>
<tr>
<td>Kansas (20)</td>
<td></td>
</tr>
<tr>
<td>Minnesota (27)</td>
<td></td>
</tr>
<tr>
<td>Missouri (29)</td>
<td></td>
</tr>
<tr>
<td>South Dakota (46)</td>
<td></td>
</tr>
</tbody>
</table>

Region 3 South

<table>
<thead>
<tr>
<th>Division 5</th>
<th>South Atlantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delaware (10)</td>
<td></td>
</tr>
<tr>
<td>District of Columbia (11)</td>
<td></td>
</tr>
<tr>
<td>Florida (12)</td>
<td></td>
</tr>
<tr>
<td>Georgia (15)</td>
<td></td>
</tr>
<tr>
<td>Maryland (16)</td>
<td></td>
</tr>
<tr>
<td>North Carolina (37)</td>
<td></td>
</tr>
<tr>
<td>South Carolina (35)</td>
<td></td>
</tr>
<tr>
<td>Virginia (51)</td>
<td></td>
</tr>
<tr>
<td>West Virginia (54)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Division 6</th>
<th>East South Central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama (01)</td>
<td></td>
</tr>
<tr>
<td>Kentucky (21)</td>
<td></td>
</tr>
<tr>
<td>Mississippi (28)</td>
<td></td>
</tr>
<tr>
<td>Tennessee (47)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Division 7</th>
<th>West South Central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkansas (03)</td>
<td></td>
</tr>
<tr>
<td>Louisiana (22)</td>
<td></td>
</tr>
<tr>
<td>Oklahoma (40)</td>
<td></td>
</tr>
<tr>
<td>Texas (48)</td>
<td></td>
</tr>
</tbody>
</table>

Region 4 West

<table>
<thead>
<tr>
<th>Division 8</th>
<th>Mountain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona (04)</td>
<td></td>
</tr>
<tr>
<td>Colorado (08)</td>
<td></td>
</tr>
<tr>
<td>Idaho (16)</td>
<td></td>
</tr>
<tr>
<td>New Mexico (35)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Division 9</th>
<th>Pacific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska (02)</td>
<td></td>
</tr>
<tr>
<td>California (06)</td>
<td></td>
</tr>
<tr>
<td>Hawaii (15)</td>
<td></td>
</tr>
<tr>
<td>Oregon (41)</td>
<td></td>
</tr>
<tr>
<td>Washington (53)</td>
<td></td>
</tr>
</tbody>
</table>

Prior to June 1984 the Midwest Region was designated as the North Central Region.