

UCLA

UCLA Electronic Theses and Dissertations

Title

Alignment between Intention and Implementation: A Case Study of the Robert Wood Johnson Foundation's Evaluation Policies

Permalink

<https://escholarship.org/uc/item/6hw4r7dg>

Author

Dillman, Lisa Marie

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Alignment between Intention and Implementation:

A Case Study of the Robert Wood Johnson Foundation's Evaluation Policies

A dissertation submitted in partial satisfaction
of the requirements for the degree Doctor of Philosophy
in Education

by

Lisa Marie Dillman

2014

© Copyright by

Lisa Marie Dillman

2014

ABSTRACT OF THE DISSERTATION

Alignment between Intention and Implementation:
A Case Study of the Robert Wood Johnson Foundation's Evaluation Policies

by

Lisa Marie Dillman

Doctor of Philosophy in Education

University of California, Los Angeles, 2012

Professor Marvin C. Alkin, Co-Chair

Professor Christina A. Christie, Co-Chair

Program evaluation plays a significant role in aiding our understanding of the effectiveness of interventions. Evaluations operate under the auspices of evaluation policies, which shape aspects of evaluation design including research questions, data collection and analysis procedures, and reporting of findings. Despite considerable speculation about the influence that evaluation policy has over evaluation practice, we do not yet have any empirical evidence to support these ideas.

Through document analysis, interviews, and an instrumental multiple case study, this study examines both the explicit and implicit policies that overarch the research and evaluation work commissioned by the Robert Wood Johnson Foundation (RWJF), the nation's largest philanthropy devoted to our public health. The primary objectives of the current study were to: 1) detail and describe RWJF's evaluation policies (including their development and communication to relevant parties); and 2) describe how the implementation of those evaluation policies happens in the field.

Findings suggest the existence of two main policy distinctions—policies that are goal statements (which describe the desired outcomes of evaluation activities) and operationalized policies (which describe the specific processes that take place during the course of evaluations). This dichotomy serves as the basis for an argument that evaluation policies function as a system, rather than existing in silos as previously described in the extant literature.

Furthermore, this study sheds light on the ways evaluation policies relate to organizational and shared learning within the Foundation. Finally, it is argued that the development of specific and directed policies to promote innovation are a necessary step to take to advance evaluation practice. Evaluation policy is one of the most critical issues facing the field of evaluation today. This study was designed to shed light on a Foundation’s evaluation policies in an effort to draw connections about how evaluation policies influence evaluation practice. The investigation helped further understanding of these connections, and generated some insight into the components of evaluation policies, and the role they might play in shaping the future of evaluation practice.

The dissertation of Lisa Marie Dillman is approved.

James W. Stigler

Louis M. Gomez

Christina A. Christie, Committee Co-Chair

Marvin C. Alkin, Committee Co-Chair

University of California, Los Angeles

2014

THIS PAGE INTENTIONALLY LEFT BLANK

To my mother, Kathy, my first teacher, demonstrating curiosity and perseverance;
my brother, David, who has shown me what grit and resilience truly are;
and my grandmother, Ruth, a model of strength and fortitude.

TABLE OF CONTENTS

LIST OF TABLES.....	IX
LIST OF FIGURES	X
ACKNOWLEDGMENTS.....	XI
VITA.....	XIV
CHAPTER 1: INTRODUCTION	1
STATEMENT OF THE PROBLEM.....	1
CONCEPTUAL FRAMEWORK	2
STUDY PURPOSE & RESEARCH QUESTIONS.....	4
STUDY SIGNIFICANCE AND IMPLICATIONS.....	5
MANUSCRIPT ORGANIZATION.....	6
CHAPTER 2: REVIEW OF RELEVANT LITERATURE	7
INTRODUCTION.....	7
EVALUATION IN THE NONPROFIT SECTOR	7
EVALUATION POLICY	11
THE CURRENT STUDY	14
CHAPTER 3: RESEARCH METHODS	15
INTRODUCTION.....	15
OVERVIEW OF THE STUDY PROCEDURES	15
STUDY SETTING.....	15
PHASE I: RECORD ANALYSIS AND INTERVIEWS.....	17
PHASE II: CASE STUDIES.....	24
ANALYTIC PROCEDURES	42
CHAPTER 4: DESCRIBING THE FOUNDATION EVALUATION POLICIES.....	49
INTRODUCTION.....	49
THE EVALUATION POLICIES OF THE ROBERT WOOD JOHNSON FOUNDATION.....	50
WHAT EVALUATION POLICIES COMMUNICATE ABOUT VARIOUS ASPECTS OF EVALUATION PRACTICE	51
EVALUATION POLICY COMMUNICATION.....	93
THE DEVELOPMENT OF EVALUATION POLICIES.....	97
CHAPTER 5: THE IMPLEMENTATION OF EVALUATION POLICIES.....	102
INTRODUCTION.....	102
THE INTERPRETATION OF EVALUATION POLICIES	102
CONDITIONS THAT SUPPORT OR INHIBIT EVALUATION POLICY IMPLEMENTATION	129

CHAPTER 6: DISCUSSION	132
INTRODUCTION	132
REVIEW OF THE FINDINGS	132
IMPLICATIONS	140
LIMITATIONS	141
DIRECTIONS FOR FUTURE RESEARCH	144
FINAL REMARKS	146
APPENDICES	146
APPENDIX A: RECRUITMENT LETTER	147
APPENDIX B: INTERVIEW CONSENT FORM	148
APPENDIX C: INTERVIEW PROTOCOL: FOUNDATION LEADERSHIP	151
APPENDIX D: INTERVIEW PROTOCOL: PROGRAM LEADERSHIP	152
APPENDIX E: INTERVIEW PROTOCOL: R&E PROGRAM OFFICER/SENIOR PROGRAM OFFICER	154
APPENDIX F: CASE STUDY: PROGRAM PERSONNEL INTERVIEW PROTOCOL	155
APPENDIX G: CASE STUDY: PROGRAM OFFICER INTERVIEW PROTOCOL	157
APPENDIX H: CASE STUDY: RESEARCH & EVALUATION OFFICER INTERVIEW PROTOCOL	159
APPENDIX I: CASE STUDY: EVALUATOR INTERVIEW PROTOCOL	161
APPENDIX J: EVALUATION POLICY CODING FRAMEWORK	163
REFERENCES.....	166

LIST OF TABLES

TABLE 1:	
RECORD DESCRIPTIVE INFORMATION	20
TABLE 2:	
INTERVIEW PARTICIPANTS.....	22
TABLE 3:	
DATA SOURCES FOR SUCCESS CASE	28
TABLE 4:	
DATA SOURCES FOR THE CHALLENGING CASE	32
TABLE 5:	
DATA SOURCES FOR COMPLEX CASE	39
TABLE 6:	
CODES NESTED IN PROCESS CATEGORIES.....	52-53
TABLE 7:	
CODES PERTAINING TO EVALUATION POLICY GOALS	54
TABLE 8:	
CODES PERTAINING TO EVALUATION POLICY ASSUMPTIONS	59
TABLE 9:	
CODES PERTAINING TO INITIATING EVALUATIONS.....	64
TABLE 10:	
CODES PERTAINING TO SELECTING AND CONTRACTING EVALUATORS	67
TABLE 11:	
CODES PERTAINING TO DESIGNING AND CONDUCTING EVALUATIONS	75
TABLE 12:	
CODES PERTAINING TO ADMINISTRATION OF EVALUATIONS.....	82
TABLE 13:	
CODES PERTAINING TO STAKEHOLDER INVOLVEMENT IN EVALUATIONS	86
TABLE 14:	
CODES PERTAINING TO REPORTING EVALUATION FINDINGS	90

LIST OF FIGURES

FIGURE 1:	
W.K. KELLOGG BASIC LOGIC MODEL.....	4
FIGURE 2:	
EVALUATION POLICY DOMAINS REPRESENTED BY EVALUATION POLICY CODING FRAMEWORK	45
FIGURE 3:	
FINAL EVALUATION POLICY CODING FRAMEWORK	46
FIGURE 4:	
PORTION OF EVALUATION COSTS COMPARED TO OVERALL BUDGET BY CASE	108

ACKNOWLEDGMENTS

My graduate school career was made possible by the unwavering faith and dedication of Marv Alkin. As my advisor for the past five years, he has provided me with innumerable opportunities both professionally and intellectually. He is a fierce advocate and phenomenally supportive. He was always generous with his praise and constructive with his criticism. I am both an evaluator and a scholar thanks to his support and I am forever grateful to have him in my corner.

Although Tina Christie was never a formal advisor, she served as a Co-Chair on this dissertation, and was a profound influence on me. Her guidance coupled with her friendship pushed me to become better than I ever thought possible. I cannot thank her enough. Jim Stigler and Louis Gomez, who rounded out this dissertation committee, provided interesting and thoughtful questions that helped shape this research. I am indebted to their participation and guidance in this project.

Incredible thanks is owed to the Robert Wood Johnson Foundation for their graciousness in hosting this study. I was constantly amazed by the participants' sharing of their time and ideas with me. They are brilliant, passionate, and committed individuals, and I am humbled by their enthusiastic participation in this project. Every conversation left me inspired, and I feel confident and excited that they are the leaders working to shape our nation's health. Particular thanks is owed to Laura Leviton, Debra Joy Pérez, Denise Herrera, Kristin Silvani, and Amy Woodrum. Laura and Debra were enthusiastic and thoughtful hosts who shared valuable insight with me and opened the Foundation's doors. Kristin, Amy, and particularly Denise provided tremendous support, logistically and otherwise.

Additional members of the SRM faculty also served key roles in my development as a graduate student and the eventual completion of this dissertation. The coursework of Fred Erickson, Reenie Webb, Felipe Martinez, and Mike Seltzer proved invaluable in the development of my methodological abilities. Terri McCarty helped hone my qualitative research skills and was a true

model for earnest and authentic qualitative research. Mike Rose in particular challenged me intellectually, and was a valuable sounding board for all of my ideas. He also helped navigate the sometimes murky waters of graduate school life and is a valued friend and confidante.

I could not have asked for a more supportive community among the other students in SRM. In particular, the other evaluation students (Nicky, Ale, Patty, Debbie, Jenn, Celina, Tim, Talia, Sebastian, and Liz), my writing group companions (Belinda, Regina, Tim, Mark, Ale, and Ji Seung), my trivia team (Jon, Belinda, and Patty) and my cohort (Scott, Jon, and Ale) have made this journey not only possible, but enjoyable. I am unendingly thankful for their support, understanding, and humor in this process! I am excited about the future with these individuals at the helm!

I particularly thank Nicky Gerardi, who has been instrumental in my development as an evaluator. Our work together over the past five years has been the most profoundly important of my graduate school career. Working with her has shaped my thinking and approach to evaluation, and will continue to do so for years to come.

UCLA Center X has also provided a valuable place for me to hone my evaluation skills. The opportunity to participate in their projects has shaped both my thinking about evaluation and my practice of it, and heavily influenced the work represented in this dissertation. I am particularly indebted to Carrie Usui-Johnson and Emma Hipolito. They are the finest teacher advocates who work tirelessly to improve education in Los Angeles schools. It has been a privilege to work with them.

My friends outside the grad school realm were also instrumental in the completion of this dissertation. Mia and Sara in particular kept me grounded and laughing at many points in this process. Nicole Eberhart was also a rock for me. We started our friendship in undergrad, and she trail-blazed this UCLA PhD journey. Over the past five years, she provided tremendous counsel, quick wit, a shoulder to cry on, and well-timed margaritas. She never doubted me, and I am so

privileged to be able to call her my friend. I also thank Danyal, Samantha, and Lisbeth, my colleagues at my “second office” for keeping me well-hydrated as I finished writing this dissertation.

I am also extremely thankful to those who helped me during the earning of my master’s degree at Temple University, particularly Will Jordan, who was a wonderful advisor and friend, and Jerry Stahler, who first introduced me to program evaluation. I also enjoyed specific help from several research assistants who contributed to the development of the coding framework in this project. In addition to their assistance validating the coding framework, Valerie, Zsanette, Shela, and Wendy helped shaped my thinking about the concepts represented here.

Lastly, I acknowledge the funding support received from UCLA’s Graduate Division, the Graduate School of Education & Information Studies (GSE&IS), and the Social Research Methodology Division within the GSE&IS while enrolled as a doctoral student at the university. The views expressed here are mine alone and do not reflect the views or policies of the funding agencies or its grantees.

VITA

EDUCATION

- 2009 M.Ed., Education
Temple University, Philadelphia, PA
- 2001 B.A., Communication
University of Pennsylvania, Philadelphia, PA

PROFESSIONAL EXPERIENCE

- 2013-2014 Teaching Assistant
Graduate School of Education and Information Studies
University of California, Los Angeles
- 2009-2013 Co-Director of Research
Teacher Initiated Inquiry Projects & Improving Teacher Quality Program
Social Research Methodology (SRM) Evaluation Group
University of California, Los Angeles
- 2009-2012 Coordinator
Graduate Student Orientation
University of California, Los Angeles
- 2005-2009 Project Coordinator
Philadelphia Bright Spaces Project
Bright Horizons Foundation for Children, Nashville, TN
- 2003-2009 Assistant Director
Kids FIRST
Bright Horizons Family Solutions, Philadelphia, PA
- 2001-2003 Infant/Toddler Lead Teacher
Kids FIRST
Bright Horizons Family Solutions, Philadelphia, PA

SELECTED PUBLICATIONS

- 2013 Dillman, L. M. Evaluator skill acquisition: Linking educational experiences to competencies. *American Journal of Evaluation*, 34(2), 270-285.
- 2013 Dillman, L. M. Comparing evaluation activities across multiple theories of practice. *Evaluation and Program Planning*, 38, 53-60.

- 2013 Wilson, C., Shepard, V., & Dillman, L. M. Is there such a thing as graduate student engagement? Yes! Case study: UCLA graduate student orientation. *NASPA Excellence in Practice: A Knowledge Communities Publication*, Fall 2012, 4-5.

SELECTED PRESENTATIONS

- 2013 Dillman, L. M. Alignment between intention and implementation: A case study of a foundation's evaluation policy. Paper presented at the 27th Annual American Evaluation Association Meeting, Washington, DC.
- 2012 Dillman, L. M. & Gerardi, N. M. Using technology to facilitate participatory evaluation: A case study of a teacher-initiated inquiry project evaluation. Paper presented at 26th Annual American Evaluation Association Meeting, Minneapolis, MN.
- 2012 Gerardi, N. M. & Dillman, L. M. The Use of a Teacher Survey to Evaluate Increases in Teacher Effectiveness in a Multi-Measure Evaluation System. Poster presented at 26th Annual American Evaluation Association Meeting, Minneapolis, MN.
- 2012 Wilson, C., Shepard, V., & Dillman, L. M. Graduate student orientation: From development to assessment. Workshop selected as the sponsored presentation by the Administrators in Graduate and Professional Student Services Knowledge Community. Session presented at NASPA: Student Affairs Administrators in Higher Education Annual Conference, Phoenix, AZ
- 2011 Dillman, L. M. Evaluator skill acquisition: Linking educational experiences to competencies. Paper presented at 25th Annual American Evaluation Association Meeting, Anaheim, CA
- 2011 Lee, C., Dillman, L. M. and Gerardi, N. Designing professional development evaluation to assess impact. Paper presented at 25th Annual American Evaluation Association Meeting, Anaheim, CA
- 2010 Dillman, L. M. Comparing evaluation activities across multiple theories of practice. Paper presented at 24th Annual American Evaluation Association Meeting, San Antonio, TX.

SELECTED HONORS AND AWARDS

- 2013-2014 Dissertation Year Fellowship, UCLA Graduate Division
- 2013 Alkin Evaluation Award (best article published in a scholarly journal); UCLA Graduate School of Education and Information Studies
- 2012-2013 Graduate Research Mentorship Fellowship, UCLA Graduate Division
- 2009-2012 William and Evelyn Hobson Fellowship, UCLA Graduate Division

CHAPTER 1

INTRODUCTION

Statement of the Problem

Significant demand is placed on service-providing organizations to demonstrate effectiveness. Program operators, recipients, funders, and the general public are all interested in knowing if an intervention works—if it is accomplishing its intended goals, and is a worthwhile investment. Myriad decisions go into addressing these issues: What questions should be asked? What information is necessary to answer those questions? Who should be responsible for collecting the relevant data? Which stakeholders should be involved in the process? How and to whom should findings be reported? The answers to these questions and more are frequently guided by evaluation policy.

Trochim (2009) defines evaluation policy as “any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation” (p.16). While it is easy to call to mind examples of formal and explicit evaluation policies, they frequently manifest in informal ways (Trochim, 2009). Most likely, an organization’s evaluation policy represents a combination of both. The knowledge about how evaluations are conducted may range widely—from tacit and unwritten to explicit and directive. This makes documenting and describing evaluation policies at an organization somewhat difficult. Regardless of the degree of formality an evaluation policy exhibits, it has the potential to play an important role both in determining an organization’s evaluation priorities, and communicating those priorities to a broad audience. More importantly, evaluation policy can greatly influence the quality of information generated through evaluations, and the usefulness of that information to inform both internal and external high-stakes decision-making about programs.

Clarifying and understanding evaluation policies within an organization is further complicated when that organization is large and hierarchical. Evaluation policies must be communicated, monitored, translated, interpreted, enforced, and enacted across a variety of offices and people. There is great potential for any policy to be added on to, subtracted from, revised, or otherwise altered (intentionally or not).

The present study addresses these issues through a case study investigation of the evaluation policies at the Robert Wood Johnson Foundation (RWJF)—a large private Foundation based on the east coast that grants money to fund interventions aimed at addressing problems related to the nation's public health. This study uses record analysis, interviews, and an instrumental, multiple case study to explore the relationship between evaluation policies and practice.

Conceptual Framework

A concept common in the evaluation literature is describing a program's theory of action—the logic that underlies the causal connections between a program's activities and desired outcomes (Alkin, 2011; Chen, 2004). For the purposes of the current study, this theory of action concept will be altered to fit a model in which the evaluation policy is substituted for a program.

An important distinction must be drawn between theory of action, and theory as it is often referred to in the social sciences. Often, rather than being based on a body of literature informed through hypothesis testing, theories of action represent hunches or notions about how a program is supposed to work (Bickman, 1987). A theory of action presents a plausible rationale that explains why the prescribed activities are expected to result in the stated outcomes. These theories may be informed in part by a body of literature, but they are often developed experientially among program staff and operators.

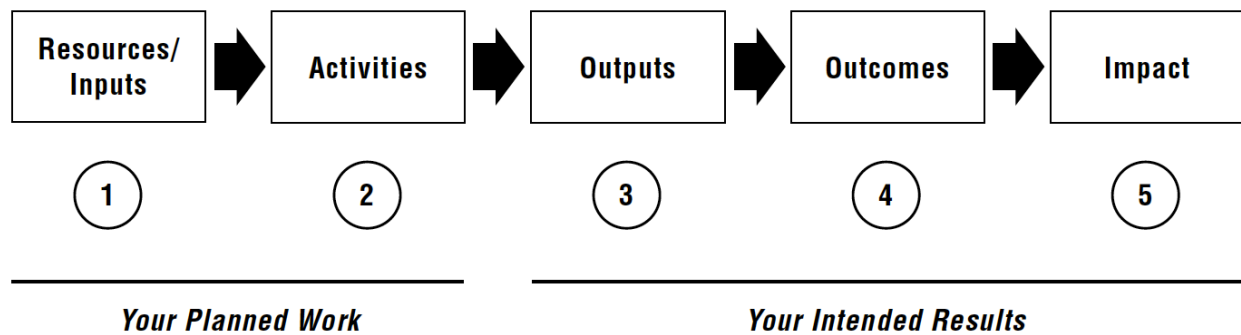
While theory of action is most commonly used when describing programs, it can be applied in this case to the evaluation policy at RWJF. The Foundation describes its evaluation policy with

particular goals in mind. While it is not necessarily expressly articulated, a theory of action underlies the choices of particular evaluative activities. In its research and evaluation framing document, the Foundation indicates that the goal of its evaluation practice is to “build an evidence base around a strategy—not to prove that our ideas work” (Robert Wood Johnson Foundation, 2010, p.1). The document also indicates that researchers will be independent, and further states, “The Foundation does not interfere with the research or evaluation it funds to alter or suppress findings in any way” (p.1). The connection between this statement about activities and the goal statement is part of a theory of action. Here, Foundation staff believe that contracting evaluations that are independent increases their objectivity, which in turn provides more reliable and convincing evidence about particular strategies.

This connection between two statements offers one example of the many possible ideas that might support the causal links between the various activities and goals articulated in RWJF’s evaluation policy documents. Making these connections explicit is an important step toward assessing the policy’s rationale (Alkin, 2011). Bickman (1987) outlines several things developing a theory of action can accomplish. The most relevant to this particular study are: clarifying the problem that is the impetus for the activities; detailing both the program components and their relative importance; and the relationship between the activities and desired outcomes (including how those might be mediated by intervening variables).

Within the context of an evaluation, a common way to depict program theory is through the development of a logic model. Logic models are graphic representations of the relationship between features common to all programs: inputs, activities, and outputs (Taylor-Powell & Henert, 2008; Wholey, 1979). An example of a basic model is presented in Figure 1.

Figure 1
W.K. Kellogg Basic Logic Model



In this model, the arrows that connect the model elements depict the theory of action. The arrows show the expected causal relationships between the elements on the model, therefore conveying the logic of how the program is supposed to work. The model is read from left to right, and can be read as a series of “*If... then...*” statements. Examining the evaluation policies in light of this framework will promote understanding of not only the particular policies, but the ways in which they interact with one another.

Study Purpose & Research Questions

The purposes of this research are two-fold: (1) to explore the evaluation policies that exist in an organization, and (2) examine the ways in which evaluation policies are implemented in real-world settings. To elaborate on these topics, the study is guided by the following research questions:

1. What are the evaluation policies of the Robert Wood Johnson Foundation?
 - a. What do the policies communicate about various aspects of evaluation practice?
 - b. Through what means are the evaluation policies communicated to those responsible for carrying them out?
 - c. How are the evaluation policies developed and determined?
2. What is the extent to which evaluations are implemented as described by the policies?
 - a. How do those responsible for enacting the evaluations interpret the evaluations?
 - b. What are the barriers to implementing evaluation policies?

- c. What are the conditions that support successful implementation of evaluation policies?

Study Significance and Implications

While the field of evaluation has engaged in considerable speculation about the importance of evaluation policy as it relates to evaluation practice, there is not yet any empirical evidence to support these ideas. To date, there has not been an intensive study about how evaluation policies are operationalized in an organization. This research investigates evaluation policy as it is described and enacted *in situ*. Specifically, it examines the creation and enactment of evaluation policy at an organization that has made a strong commitment to evaluation, and has developed and articulated an evaluation policy. It provides an empirical compliment to the theoretical work that has been done thus far about evaluation policy. The research has both local and global implications. First, and most directly, it informs the Foundation where the research took place about the fidelity with which its policies are implemented. More broadly, it speaks to evaluation efforts of nonprofits. Finally, it contributes empirically to the conversations generally about evaluation policies.

This study has immediate relevance to RWJF, where the research was conducted. According to RWJF's President and CEO, a primary goal of its research and evaluation efforts is to “produc[e] evidence that both policy makers and practitioners in the field can use” (Robert Wood Johnson Foundation, 2012). Significant resources are devoted in this regard; approximately 20% of the monies spent on grant allocations are dedicated to research and evaluation efforts (Robert Wood Johnson Foundation, 2010). Furthermore, RWJF has been intentional and deliberate about developing evaluation policies. From the perspective of the Foundation, the information presented in this study helps to answer underlying questions about the value and results of having such policies. These answers may be transferable to other foundations, nonprofit organizations and/or public agencies that conduct or contract evaluations.

Beyond organizations and agencies that commission evaluation, this study sheds light on the ways in which evaluation policies interact with one another and relate to evaluation activities. Furthermore, the ways in which evaluation policies contribute to organizational learning and the practice of evaluation more generally are explored.

Manuscript Organization

There are five chapters that make up the remainder of this manuscript. In Chapter 2, a review of the relevant literature that helps frame this investigation is presented. Chapter 3 offers information about the study's methods and procedures for collecting data as well as the analysis procedures. The study findings appear across two chapters, each of which answers one of the main research questions. Chapter 4 includes information describing and detailing the Foundation's evaluation policies and Chapter 5 describes the implementation of those policies in practice. In the final Chapter (6), conclusions from the analysis of this study in addition to its implications, limitations, and possible directions for future research are shared.

CHAPTER 2

REVIEW OF RELEVANT LITERATURE

Introduction

Evaluation policy has only recently been explored by evaluation scholars (Mark, Cooksy, & Trochim, 2009). Despite its relatively new emergence in the literature, it is touted as one of the most critical issues facing the field today (Datta, 2009; Mark et al., 2009; Trochim, 2009). Discussions of evaluation policy are often undergirded by federal legislation aimed at increasing the accountability of government programs and demonstrating program effects. These principles have spread into the nonprofit sector, which struggles to keep up with growing evaluation demands. Private Foundations, because of their perceived potential to advance knowledge, experience pressure to demonstrate value (Porter & Kramer, 1999).

Evaluation in the Nonprofit Sector

Evaluation in the nonprofit sector has evolved over the years against a backdrop of federal legislation aimed at increasing accountability and demonstrating effectiveness. Nonprofits over the past several decades have made strides to become more systematic about describing the effects of what they do, but there is still significant work to be done (Flynn & Hodgkinson, 2001).

Private foundations, such as the one that is the subject of this investigation, comprise one of the 30 different distinctions of nonprofit status according to the U.S. tax code. There are an estimated 2.3 million nonprofit organizations operating in the U.S.¹, of which nearly 99,000 are private foundations (National Center for Charitable Statistics, 2012). Nonprofits are incredibly diverse in scope of services provided (some examples include arts, culture, and humanities;

¹ This figure is approximate because it includes an estimate of the number of organizations that do not have to register with the IRS (those with less than \$5,000 in yearly revenue and religious congregations and their auxiliary groups). In 2010, 1.6 million nonprofit organizations registered with the IRS.

education and research; health services; environment and animals; and religion). Despite the wide range of nonprofit organizations, they share some common characteristics, including:

- They are formally organized with a legal identity
- They are private (separate from government)
- If profits are generated, they must be redistributed to the mission of the organization (rather than distributed to the organization's founders)
- They are self-governing (internally managed)
- They involve some degree of voluntary participation (e.g. a voluntary board of directors, or volunteer corps that are responsible for carrying out activities)
- They are of benefit to the public (National Center for Charitable Statistics, 2012)

This last point is particularly important. Organizations in the nonprofit sector are designed to provide some sort of public good or service, often to fill in gaps left by the inherent limitations of the government and the market as such providers (Salamon, 2001). The pace at which government is able to mobilize in response to a need can sometimes be extremely slow due to the necessity of reaching a majority in support of any action. Furthermore, even when a majority is reached, the bureaucracy that may accompany the action could prove too cumbersome for effective operations. The market, on the other hand, while unencumbered by some of the operational struggles associated with government, is best suited to providing individual goods (e.g. shoes) rather than collective ones (e.g. clean air) (Salamon, 2001). Nonprofits, therefore, are able to mobilize common interests quickly and to meet public needs unattended to by government or the free market.

Foundations play an intermediary role in this system. The majority of private foundations do not provide direct services; rather they support (financially and otherwise) the efforts of other

nonprofits². In 2012, foundations gave approximately \$50.9 billion in grants to other organizations (The Foundation Center, 2013). Prior to the 1980s, one of the primary perceived benefits of foundations was that they served as a breeding ground for new ideas with flexibility to experiment with new approaches (Behrens & Kelly, 2008). In order for this to be effective—for foundations to be able to disseminate accurate information about best practices to a wide audience—they need a somewhat sophisticated system for determining which models work and which do not.

The nonprofit sector in general faces external pressure from a variety of sources to demonstrate effectiveness of their operations. While many people agree that nonprofits play a critical role in providing necessary public services that otherwise could not be delivered, the public has very little data that speaks to the results of the programs and interventions (Flynn & Hodgkinson, 2001). This is concerning considering that nonprofits contribute over 800 billion to the U.S. economy yearly (5.5% of the country's Gross Domestic Product) and account for a little over 9% of all wages and salaries paid³ (Roeger, Blackwood, & Pettijohn, 2012). The general measure of success for any organization is whether or not it creates more value than it consumes. However, unlike in the for-profit realm, which can count profits and losses in dollar amounts, the value created through nonprofit organizations is benefit to the public. This is much more difficult to measure—benefits may be intangible, and any effects might be distal and/or widely dispersed (Oberfield & Dees, 1992).

The expectation that nonprofits demonstrate effectiveness has coincided with the passage of federal legislation aimed at similar ends for government-run programs (Carman, Fredericks, & Introcaso, 2008). Notably the National Performance Review and the Government Performance and Results Act (GPRA) of 1993 were designed to limit bureaucracy and increase accountability among

² An exception to this is the small number of operating Foundations that provide direct services.

³ Data is from 2010.

government programs. In 2004, the Bush administration developed and started promoting the Program Assessment Rating Tool (PART)—a questionnaire designed to assess and improve program performance (Carman et al., 2008). These tools and legislative actions are a part of an era of accountability that has seeped into all parts of the public sector.

Despite increasing calls for accountability, transparency, and demonstrating effectiveness, many foundations have been slow to adopt evaluation practices. In 1990, few foundations demonstrated understanding of evaluation methods, let alone use of them (Behrens & Kelly, 2008). This has changed in the intervening years—a 2004 study indicated that 31% of small foundations and 88% of larger foundations engage in some sort of formal evaluation process (Ostrower, 2004). Despite increased attention paid to evaluation, foundations are criticized (largely from trustees and executives) for the lack of useful evaluation results (Behrens & Kelly, 2008).

These criticisms are partially magnified because of the role foundations play as brokers of funds to those that provide social services. Porter and Kramer (1999) argue that to compensate for their position as expensive middle men, foundations should work to create benefit in ways that extend beyond their purchasing power including: (1) selecting the best grantees; (2) signaling other funders; (3) improving the performance of grant recipients; and (4) advancing the state of knowledge and practice (p. 123-125).

The third and fourth points are directly related to evaluation. In order to act on either, a foundation needs reliable information about the performance of interventions in the field. This information is gathered through evaluations. For those foundations that do participate in formal evaluations, these investigations are likely not haphazard. Studies are often commissioned through a request for proposal (RFP) process, which formalizes many elements of the evaluation (Alkin, 2011). Often, what is communicated in an RFP, deliberately or not, is information that reflects an evaluation policy. RFPs may signal preferred methods, levels of stakeholder involvement, available

resources, and information about project management. All of these things will impact the type and quality of data received, and are informed by an evaluation policy (Trochim, 2009). This policy, however, might be designed or unintentional, and may be understood, rather than explicitly articulated. The idea of evaluation policy is a relatively new, but important area of research. Evaluation policies are potentially very powerful, but all too often, too little attention is paid to them (Mark et al., 2009).

Evaluation Policy

An important distinction must be made between evaluation policies and broader substantive policies. An example of the latter (drawn from education) is the No Child Left Behind Act of 2001, which sought to close the achievement gap in schools and ensure that all children reach proficiency levels by 2014 in basic subjects taught in K-12 education (*No Child Left Behind Act of 2001*, 2002). This is a policy because it has a definable outcome, clear time frame, and offers a general statement, but lacks specifics, about how the goal will be achieved (Trochim, 2009). Evaluation policy, on the other hand, has been defined as “... any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation” (Trochim, 2009, p. 16). This is not to say that evaluation policy and substantive policies are unrelated. Evaluation policies ultimately shape evaluation practice, which influence programs and interventions, and subsequently the substantive policies that overarch them (Trochim, 2009).

The definition that Trochim offers leaves room for both implicit and informal guidance, and explicit and formal guidance about performing evaluations (Mark et al., 2009). Regardless of how specifically and intentionally they are articulated, every group that engages in evaluation has evaluation policies, and all evaluations are in some way influenced by those policies (Trochim, 2009).

Given that evaluation policy, even when invisible, guides how all evaluations are conducted, it is of critical importance, both “enabling and constraining the potential contributions evaluations

can make” (Mark et al., 2009, p.3). It can communicate what an organization values in terms of evaluation. It can signal the amount of resources devoted to evaluation efforts, the theoretical predispositions toward evaluation, and how evaluation findings should be used (Trochim, 2009). Furthermore, when evaluation policies are explicitly articulated, they can be subjected to a democratic process. When something is stated, it can be contradicted and challenged, or expanded and elaborated (Trochim, 2009). A dialogue can ensue, the end result of which might be a stronger policy.

Trochim (2009) further argues that evaluation policy is the most important matter facing the field of evaluation today because it encompasses all other aspects of evaluation practice that are often topics of debate among practitioners and scholars (methods, stakeholder involvement, use, dissemination of information, validity, etc.). Beyond the implications for the field, evaluation policies have direct impact on what social programs are funded, and their perceived value (Datta, 2009).

As a complement to the broad considerations of defining evaluation policy and describing its importance, there has been some consideration of the specific components that might make up an evaluation policy. Mark et al. (2009) point out that while methods are the most obvious candidate for evaluation policy elaboration, evaluation policy can describe myriad evaluation features. Some of these include funding, contracting, evaluator qualifications, stakeholder involvement, contextual considerations, dissemination of findings, and use.

Some work has been done in the field to further explicate the idea of evaluation policy, and discuss it in terms of its specific aims and components. Specifically, in 2007 the American Evaluation Association commissioned a task force to focus on issues of evaluation policy (American Evaluation Association, 2007), and Trochim (2009) contributed a framework for describing evaluation policy.

The American Evaluation Association's Evaluation Policy Task Force (EPTF) outlines seven areas of evaluation policy:

- **Evaluation definition.** How, if at all, is evaluation defined in an agency or in legislation? In such contexts, how is evaluation formally distinguished from or related to other functions such as program planning, monitoring, performance measurement or implementation?
- **Requirements of evaluation.** When are evaluations required? What programs or entities are required to have evaluations? How often are evaluations scheduled? What procedures are used to determine when or whether evaluation takes place?
- **Evaluation methods.** What approaches or methods of evaluation are recommended or required by legislation or regulation, for what types of programs or initiatives?
- **Human resources regarding evaluation.** What requirements exist for people who conduct evaluations? What types of training, experience or background are required?
- **Evaluation budgets.** What are the standards for budgeting for evaluation work?
- **Evaluation implementation.** What types of evaluation implementation issues are guided by policies? For instance, when are internal versus external evaluations required and how are these defined?
- **Evaluation ethics.** What are the policies for addressing ethical issues in evaluation?
(American Evaluation Association, 2007)

Trochim's (2009) evaluation policy framework includes eight complementary, yet slightly different elements than those presented by the EPTF. These include: (1) evaluation goals and policies that indicate the desired outcomes of evaluations; (2) evaluation participation policies that detail who is involved in evaluations, when, and under what circumstances; (3) evaluation capacity building policies that describe the intention of the organization to build its capacity to conduct

evaluations; (4) evaluation management policies that inform the allocation of resources (financial and otherwise) and oversight; (5) evaluation roles policies that outline the varying responsibilities of actors involved in conducting evaluations; (6) evaluation process and methods policies that govern the procedures of conducting evaluations; (7) evaluation use policies that describe how evaluation results will be communicated and acted upon; and (8) meta-evaluation policies that indicate how organizations might periodically assess the quality of their evaluations.

These two frameworks were generated from a theoretical, rather than empirical standpoint. The frameworks played an essential role in informing the analysis for the present study. In addition to exploring how the observed evaluation policies of RWJF map onto these existing frameworks, this study also explores the ways in which these domains interact with one another—a factor that is not present in the extant discussions about evaluation policies.

The Current Study

In sum, despite assertions that evaluation policy is a critical issue facing the field today, there has been no empirical investigation of these issues to date. Thus far, the scholarship in this area consists of theoretical explorations of evaluation policies and well-informed speculations about the factors of evaluation practice that evaluation policies may influence. The present study provides more of an in-depth, field-based exploration of evaluation policies both in terms of their content, and the ways in which they interact with one another. Furthermore, the implementation of these policies during evaluations is examined. These issues are explored through a qualitative study including record analysis, interviews, and case studies. Data were analyzed such that the theories of evaluation policies detailed here can be expanded. The UCLA Institutional Review Board (IRB) approved all data collection activities (UCLA IRB# 13-00456).

CHAPTER 3

RESEARCH METHODS

Introduction

This chapter presents the methods used to address the research questions of this study: What are the evaluation policies of the Robert Wood Johnson Foundation? What do the policies communicate about stakeholder involvement, evaluation methods, evaluation goals, etc.? Through what means are the evaluation policies communicated to those responsible for carrying them out? How are the evaluation policies developed and determined? What is the extent to which evaluations are implemented as described by the policy? How do those responsible for enacting the evaluations interpret the policies? What are the barriers to implementing evaluation policies? What are the conditions that support successful implementation of evaluation policies? This chapter begins with an overview of the study procedures. Then, the setting for the study, the participants and instruments, and specific data collection procedures are described. Finally, the analytic procedures are explained.

Overview of the Study Procedures

To address the research questions, this study was conducted in two stages. First, record analysis and interviews were conducted to explore the first set of research questions about developing an understanding of the evaluation policies of the Robert Wood Johnson Foundation. These data collection activities then informed the second phase of the study: an instrumental multiple case study designed to explore the implementation of those policies in real world settings.

Study Setting

This study was conducted at the Robert Wood Johnson Foundation (RWJF) located in Princeton, New Jersey. The Foundation has been in operation for more than 40 years, and focuses

on health issues in the U.S. and U.S. territories. The Foundation's assets place it among the top 10 Foundations in the United States, and it is the largest philanthropy dedicated solely to the nation's health. RWJF has been long committed to evaluation efforts and is considered a leader among its peer organizations (Hall, 2004). The Foundation specifically articulates many broad policies that govern its work in evaluation through specific framing documents, instructional materials, and webpages. Furthermore, RWJF dedicates significant funds—approximately twenty percent of its granted dollars—to its research and evaluation efforts.

Foundation Organization

The Foundation is structured in a way that supports reaching several strategic objectives. The main mission of the Foundation (to improve public health) has remained the same over its decades-long history; however, the specific areas of interest and mechanisms identified for change have shifted over time as the context and Foundation leadership undergo changes. During the time in which this study was conducted (starting in the spring of 2013), the foundation was undergoing a comprehensive strategic planning effort, which somewhat altered its structure. This description pertains to the organization and operations that were in place at the beginning of data collection processes.

The Foundation uses a matrix model of management, under the direction of a Board of Trustees. There are two primary distinctions upon which the Foundation derives its organizational structure. The first is programmatic and the second is by task. Programmatically, there are two primary groups that oversee the work of seven teams, each with a specific strategic area of focus. The tasks of the Foundation are divided into departments, some of which primarily function along program lines (for example, research and evaluation, and communications) and others provide a broader base of general Foundation support (for example, human resources and information technology).

Of interest to this study is the programmatic work. At the time of this study, the Foundation had set objectives and indicators for each of its seven targeted teams; those objectives and indicators were used to measure its progress. The teams fund specific programs that vary widely in size, scope, and objective, but that the foundation believes will contribute to moving these indicators. Each funded program is overseen by a program staff team that includes one program officer, one evaluation officer, and one communications officer that work together. As this study focuses on the evaluation policies of the foundation, the primary unit of focus is the Research and Evaluation (R&E) department. R&E officers are assigned to work on multiple programs and as such, their work may cross multiple teams.

Phase I: Record Analysis and Interviews

The first set of research questions seek to define and describe the evaluation policies of RWJF and the various aspects of evaluation practice that the policies overarch. Many of these questions can be answered through analysis of records, both public and internal, written by and about the Foundation. However, evaluation policies can exist both formally and informally, be tacit or explicit, articulated or inherited. Complexities arise in trying to illuminate those policies that are not explicitly written, and in locating their origins. In these cases, the analysis of RWJF records was supplemented by interviews of key Foundation staff in the Research and Evaluation Unit.

The next research question seeks to understand how these policies are communicated by RWJF. Evaluation policies are intended to translate into action, but the journey between the paper and the field has many obstacles, one of which is the communication of the policies among the various actors in the system. The Foundation makes deliberate choices about the message it sends about evaluation policy, and the means through which that message is delivered. The record analysis and interviews in this phase of the investigation help illuminate the ways in which the evaluation policies are communicated.

Record Analysis

RWJF creates a wide array of records⁴, both publicly and internally that describe its philosophies, activities, and accomplishments. These take the form of reports, meeting minutes, anthologies, framing documents, videos, interviews, etc. Beyond those items that the Foundation produces and maintains on its own, there are numerous publications in academic and practitioner journals, press coverage items, and studies conducted externally about Foundation practices.

The journey that policies take through inception, articulation, interpretation and enactment might not be linear or intentional. While some policies are clearly developed in top-down fashion, it is conceivable that some emerge from best practices discovered in the field. Furthermore, while evidence of evaluation policies will appear in documents designed for that express purpose, it might also be embedded in documents seemingly unrelated to evaluation, and be discussed in internal communications. There are several possible trajectories an evaluation policy may follow, and multiple paths through which it might be expressed. To ensure sufficient exploration of the available avenues for discovering evidence of evaluation policy, two different techniques to construct the sample and complete the analysis were employed. The first sampling technique is a top-down approach that attempts to gather the comprehensive and official description of the Evaluation policies of RWJF; the second technique used was staff interviews.

Top-down sampling. The collection of documents for review and analysis began with the Research and Evaluation framing document (Robert Wood Johnson Foundation, 2010). This is the primary document used to communicate information about the research and evaluation policies at the Foundation. Included in the document are references to other documents and sources relevant to the Foundation's evaluation policy that were subsequently added to the sample. This procedure can be considered a form of snowball sampling, the traditional conception of which involves

⁴ The term record is used, rather than the more familiar term, document, because the items in this selection refer to a variety of media including videos.

accessing informants through contact information provided by other informants (Noy, 2008). For the purposes of this study, rather than informants, the origin of the sample is the research and evaluation framing document. The references contained therein served as second-level sources.

Interview sampling. Documents were also added to the sample based on conversations that occurred over the course of the interviews. Key RWJF personnel were asked to provide examples of evaluations that they considered to be particularly useful for the Foundation's aims, and those they felt fell short of the Foundation's standards. These exemplars were added to the sample. Including these types of documents affords an opportunity to examine the relationship between R&E staff perceptions of quality to the consistency with which evaluations adhere to policies.

Description of Foundation evaluation policy records. The pool of records generated through the above-described procedures was extensive. Different sources assisted in informing different aspects of the study. Many of the records were useful in facilitating contextual understanding of the Foundation's structure and practices. Ultimately, some sources or records did not prove useful in enhancing understanding of the Foundation's evaluation policies. For this reason, the sample of records used for analysis was ultimately restricted to those produced by the Foundation that specifically include explicit articulations of evaluation policies based on preliminary coding (Table 1). The titles of the records are not included in the table; rather, they are lettered A through T. The final sample includes twelve videos and eight documents.

Table I
Record Descriptive Information

Record	Excerpts	Codes	Audience	Type	Topic
A	112	158	Evaluators	Document	Stakeholders
B	63	109	Public	Document	Operations
C	34	46	Public	Document	Policy
D	17	22	Public	Document	Policy
E	12	15	Grantees	Document	Policy
F	8	9	Internal	Video	Policy
G	8	10	Internal	Video	Policy
H	7	7	Internal	Video	Policy
I	7	8	Evaluators	Document	Reporting
J	6	6	Internal	Video	Operations
K	4	3	Internal	Video	Operations
L	4	7	Internal	Video	Policy
M	4	5	Internal	Video	Policy
N	3	5	Internal	Document	Operations
O	3	2	Internal	Video	Operations
P	3	3	Internal	Video	Operations
Q	3	4	Internal	Video	Reporting
R	2	1	Internal	Video	Policy
S	1	4	Internal	Document	Operations
T	1	1	Internal	Video	Operations
Total:	302	425			

The table indicates the number of selected excerpts and codes applied to each document. For the purposes here, a code is “a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data” (Saldaña, 2009, p. 3). The specific coding scheme is discussed later in this chapter (p. 42). The numbers in the columns differ because in some cases several codes were applied to the same excerpt (425 codes are applied to 302 excerpts). The table shows that some topic areas are more represented in the coding scheme than others. The greatest number of codes appear in the one document that pertains to stakeholders (37% of excerpts, and 37% of codes overall). This is followed by policy documents, which include 32% of excerpts and 29% of codes. The category of operations closely follows these totals with 28% of the excerpts and 31% of the codes. Finally, records that reference

reporting contain three percent of the excerpts and three percent of the codes. The differential application of codes according to topical area is explored further in the chapter that follows.

Interviews

Interviews play a key role in furthering understanding of the evaluation policies of the Foundation and the ways in which they are developed and communicated. Interviews serve as an excellent complement to the record review because they help illuminate evaluation policies that are held more informally and not articulated as explicitly. Two key RWJF informants helped facilitate the interview process by: providing general background knowledge of the Foundation, determining who best to interview, refining interview questions, and scheduling staff interviews. Each of these processes is described further below.

Participants. Initial interviews occurred with several members of the Research and Evaluation staff at RWJF, and with Foundation leadership members in-person at the Foundation headquarters. At the time this study began, there were 16 staff members working in the Research & Evaluation Unit. Interview requests were sent to all 16 R&E staff via the key Foundation informants using the Foundation's internal calendar and meeting request system. An introductory letter, co-authored by the researcher and one of the key informants, and a consent form were sent to the prospective participants to explain to them the nature and scope of the investigation (Appendices A and B). The content of the consent form was informed by guidelines and requirements mandated by the Institutional Review Board at the University of California, Los Angeles. Signed consent forms were obtained via email prior to the start of each interview.

Of the 16 individuals from the R&E unit contacted for an interview, 13 participated. In the other three cases, conflicts prohibited scheduling of an in-person interview. In addition to the R&E staff, requests for interviews were sent to five Foundation leadership members, all of whom participated in an interview. The majority of the interviews were conducted in May of 2013 at the

Foundation offices. Three follow up interviews took place at the Foundation in March of 2014.

Table 2 displays basic information about the interview participants. Each interview ranged in duration from 30 minutes to an hour. In some cases, participants were interviewed more than once.

The total time spent speaking with each participant is indicated in the interview duration column. All interviews were conducted in person with the exception of the interview with Participant 6, which was conducted over the phone.

Table 2
Interview Participants

Participant	Department	Title	Foundation Tenure (Years)	Interview Duration (Minutes)
1	Foundation	Chief of Staff	14	60
2	Foundation	President and CEO	10	30
3	Program Leadership	Associate Vice President	16	30
4	Program Leadership	Assistant Vice President	10	60
5	Program Services	Services Coordinator	12	30
6	R&E	Vice President	15	60
7	R&E	Web Editor	13	30
8	R&E	Assistant Vice President	9	60
9	R&E	Program Officer	2	30
10	R&E	Senior Program Officer	11	30
11	R&E	Senior Advisor	14	150
12	R&E	Senior Program Officer	5	60
13	R&E	Program Officer	1	60
14	R&E	Senior Officer	15	120
15	R&E	Senior Program Officer	16	30
16	R&E	Research Assistant	1	60
17	R&E	Senior Program Officer	2	60
18	R&E	Senior Program Officer	1	45

Most participants were R&E program officers and general staff. Several other interviews were conducted with Foundation leadership, including the CEO, Chief of Staff, and the Associate Vice Presidents responsible for the two major strategic initiatives of the Foundation. For information gathering purposes, interviews were conducted with non-R&E staff; those interviews were not

included in this part of the analysis. Participants had widely varying years of experience at the Foundation, ranging from just under 1 year to nearly 17.

Procedures. Interviews were semi-structured, and explored topics that spanned the majority of the aforementioned research questions. Each interview began with a review of the consent form. Protocols differed slightly based on the participant's job within the Foundation. For those with more leadership responsibilities (Appendices C and D), interview topics covered elements of the evaluation policies, the history of how they are developed and determined, the communication strategy for policies, evaluation report exemplars, and programmatic changes and advocacy efforts that have been informed by the results of evaluations. Some sample questions about broad RWJF functions as they relate to evaluation policy include: Who is responsible for determining evaluation policy at RWJF? In what ways is evaluation policy communicated? When an evaluation is contracted by the Foundation, how do staff members in the Research and Evaluation unit ensure that the evaluation is carried out in ways consistent with the Foundations policies? Will you share an example of when this worked particularly well? Will you offer an example of an evaluation report that did not meet the Foundation's standards for evaluation quality? How do you determine what information to share and with whom to share it? How often are the evaluation policies revisited? What factors influence any policy changes that are made?

For R&E program officers who directly oversee evaluation activities, questions included information about the processes of supervising evaluation work in addition to some of the higher-level questions described above (the full protocol can be found in Appendix E). Specific questions included: How do you decide whom to contract for an evaluation? What is the process for determining research questions? How do you identify key stakeholders? How are stakeholders engaged in the evaluation process? How often and in what ways do you communicate with

evaluators during an evaluation? How do you decide on final deliverables from a project? In what ways are evaluation findings used?

All interview participants were offered the interview questions in advance if they wished to be afforded time to prepare their responses. The same interview protocol was used for each interviewee, but the questions were slightly tweaked based on participant experience.

Follow-up in-person interviews occurred in March of 2014. These interviews were informed by preliminary analysis of the records and first-round interviews, and primarily served a clarification purpose. Furthermore, the interviews helped ensure data saturation. There are two considerations to take into account for data saturation: appropriateness and adequacy (Fossey, Harvey, McDermott, & Davidson, 2002). Appropriateness was assured given the ability to interview a significant portion of the small number of potential interviewees. The second round of interviews afforded opportunities for discovering recurring patterns of themes, thus meeting the criteria of adequacy (Fossey et al., 2002).

Phase II: Case Studies

The second set of research questions explores the implementation of the evaluation policies. The alignment of conducted evaluations to the intentions outlined in the policies provides an interesting starting point for investigating the conditions that support successful policy enactment, and the barriers to proper implementation.

While RWJF makes specific and deliberate efforts to explicitly communicate its policies to evaluators, the policies are flexible enough that evaluators have room to adapt evaluations to meet the needs of particular contexts and situations. This part of the research explores how evaluators fill in the blanks, so to speak. Aside from what is explicitly written, R&E program officers and evaluators make inferences about the intentions of the policies, and develop proposals believed to best match the interpretation of those policies. Alternately, evaluators may disagree with the dictates

of a policy and carry out an evaluation that does not exactly match what is prescribed.

An instrumental, multiple case study is employed to explore these questions. Three RWJF-funded evaluations were selected to explore the implementation of evaluation policies. According to Stake (1995), a case study is instrumental when it is used in order to further understanding of something else. In this study, the cases of Foundation-funded evaluations are used to reflect back on the practices of the Foundation as a whole.

Case Studies

Three RWJF funded evaluations were selected for close examination to address the second set of research questions about the field implementation of the evaluation policies. These cases were purposefully sampled (Patton, 2002) with assistance from foundation staff. To facilitate comparisons, a successful evaluation (Case 1) and a challenging evaluation (Case 2) were selected. The third case was selected based on the assumption that because of its prominence, it is the best representation of evaluation policies in the field given that it is one of the most expansive, complex, and resource-demanding programs and evaluations currently operating under the auspices of the foundation. This section summarizes the selection of cases and the case study procedures and includes a descriptive write-up of each case.

Sample of Cases

Case selection was facilitated through the key Foundation informants and initial interviews. Each interview participant (from the original interviews in May of 2013) was asked to share examples of evaluations that they considered to have gone well, and others that they thought were not successful. Key informants were consulted to gather more information about these particular programs and their attendant evaluations. Ultimately, cases were selected as the focus of this part of the investigation based on consensus of their strengths and challenges according to Foundation personnel, and the availability of information about each case.

Procedures

Each case was explored through a combination of document analysis and interviews. According to Yin (2009), these data sources represent two of the most commonly selected for case studies. Each source has strengths and weaknesses associated with its inclusion in a case study. Documents are the most commonly included data source for case studies, considered desirable because they are stable, unobtrusive, exact, and broad. The potential drawbacks associated with using documents most relevant to this study include biased selectivity and reporting bias (from the author) (Yin, 2009). However, neither of these is significantly detrimental to the study at hand. Even though not every document produced as a part of these evaluations was incorporated into the analysis, their selection represented all phases of the evaluation. The potential for reporting bias is also limited because many of the documents were subject to a peer-review process, or a less-formal, but still rigorous vetting process within the Foundation.

Interviews are commonly included in case study research because they are considered targeted and insightful means to gather information. Inaccuracies due to poor recall are of a concern in this study because in some cases, participants were asked to answer questions about things that happened several years ago. For this reason, interview protocols were shared with participants in advance. Many participants indicated through the conversations that they had reviewed notes and files in preparation for the calls.

Document sample. Documents were selected throughout the history of each case through a combination of foundation assistance (for the internal documents) and web searches. The foundation maintains an internal program information and management system, which tracks each of its funded programs and the records produced under its auspices. The main documents selected for inclusion in this case were drawn from this list. These include requests for proposals, responses

to those requests, contracts or agreements for the scope of work (called a *précis* in the language of the foundation), interim and final reports, and published products for the foundation.

Interview participants. Interviews were conducted with key parties associated with each evaluation: Foundation program staff (i.e., those who managed the program being evaluated), R&E staff (i.e., those responsible for overseeing the evaluation) and external evaluators (contracted by the Foundation to conduct the evaluations). Additionally, some Foundation leadership personnel were consulted for their opinions and experiences. Finally, program operators were asked to participate. Initial recruitment emails were sent by the Foundation, and they included the study consent form, as per UCLA IRB's standard procedure. Interview protocols were developed based on the research questions of interest, and the findings from the first phase of analysis. Protocols differed slightly based on the particular role that the person had in the evaluation. Appendix F is the full interview protocol for program personnel; Appendix G is the interview protocol for the Foundation program officer; and Appendices H and I share the interview protocols for the Foundation R&E Officers and the evaluator, respectively.

With few exceptions, all interviews were 60-minutes in length and were conducted via telephone. Interview protocols were emailed to participants two days in advance of the scheduled interview. The next section of this document includes descriptive write-ups of each case. The number of documents and respondents is included in each of the case write-ups.

Case 1: Success case. This first case was selected with assistance from the Foundation as an example of what it considers to be an exemplary evaluation. This program was subject to two separate evaluations. The first was an implementation study, which was followed a year later by an impact study. It was described as having a rigorous design, but the Foundation noted such a design is “actually a departure from our usual practice.” Table 3 shows the data sources consulted in developing this case study. This section includes a brief overall summary of the history and

background of this program⁵ followed by an overview of the evaluation activities associated with it. Finally, the processes of both evaluations are presented more in-depth.

Table 3
Data Sources for Success Case

Documents	
	Program theory of change
	Implementation study proposal
	Implementation study contract summary
	Implementation study annual narrative
	Implementation study final report
	Implementation study final narrative
	Peer-reviewed article based on implementation study
	Impact study successful proposal
	Impact study unsuccessful proposal
	Impact study contract agreement
	Impact study annual narratives (years 1-4)
	Impact study executive summary
	Impact study issue brief
Interviews	
	Impact study lead evaluator
	Implementation study lead evaluator
	Foundation program officer
	Foundation research and evaluation program officer
	Program founder
	Foundation senior vice president

Program history and description. The Fun and Games program is a school-based initiative designed to positively affect student- and school-level outcomes. It is a sports program delivered by coaches that focuses on inclusive play. The program was founded in 1996 in response to a challenge the founder observed at school sites. Essentially, the founder thought there was a way to make time outside the classroom more productive for students and schools. The founder developed this program, and implemented it in two schools in New Jersey City. Ultimately, the program sought to increase student engagement and improve school climate. During the next several years, the program expanded within the initial community, and by 2003, the program had reached several additional schools in a neighboring city.

⁵ All of the names and some identifying details of this and all the other cases have been redacted and/or changed for confidentiality purposes.

It was around this time when the Foundation became aware of the Fun and Games program. A Foundation program officer heard about Fun and Games from a trusted and respected colleague while attending a conference. The officer became intrigued and sent a consultant to observe the program. The consultant highly recommended the program, and the Foundation program officer approached the founder with a question, “what would [you] want to do with funding from a Foundation?” The founder’s initial reaction was to request \$200,000 to write a book. This was not exactly what the Foundation program officer had in mind.

At that time, the Foundation was developing a model whereby it supported its strategic initiatives through the identification of a small, seemingly successful program, “developing its evidence base and then scaling it nationwide.” Rather than funding the writing of a book, the Foundation program officer was more interested in applying this scaling strategy to Fun and Games. After more discussion and back-and-forth, it was eventually agreed that the program would receive several million dollars to replicate in 4 different cities nationwide in the following three years.

This demand for replication proved to be an “upheaval” for the program. The shift necessitated dramatic staffing expansion, which proved challenging for the Fun and Games program. These challenges led to a situation in which the program ran considerably over budget due to problems with timely payment from the schools where the program was operating. The Foundation struggled to support them through this time because “we [the Foundation] didn’t have many grantees in this situation where they actually have a revenue stream coming in to help them grow ... We [the Foundation] were evolving our strategy of scaling and hadn’t really factored in cash flow would be an issue for scaling at least a partially self-sustaining model.”

Eventually, the issues were resolved, and Fun and Games became more adept at operating on a larger scale. During this time period, the program was engaging in self-evaluation that relied heavily on satisfaction surveys and anecdotal evidence. The Foundation had not yet attached an

evaluation to this particular project. After the initial 3-year period, the Foundation decided to expand the program to 27 large- and mid-sized cities through a considerable grant. It was at this point that the program became a significant enough investment to warrant an evaluation.

Program evaluations. Initially, the Foundation was interested in conducting a randomized control trial (RCT). A Research and Evaluation program officer at the Foundation approached colleagues at a research center with close ties to a prestigious research university to discuss their potential involvement in the project. RCTs are not, however, part of the work done at this particular research center, so the center “sort of convinced [the Foundation] that before it could even start a randomized control trial, it needed to do an implementation study because [the Foundation doesn’t] have any idea even what it should be looking for in terms of outcomes unless it understands how the program is implemented.”

After several conversations between the R&E program officer and the implementation evaluator at the research center, an agreement was reached whereby the center would conduct an implementation study across one academic school year. This mixed-methods study relied on “interviews, observations, teacher and student surveys and focus groups, and reviews of school records” to assess and compare the implementation procedures at 8 schools that were newly implementing the program, and 4 schools with more extensive Fun and Games experience. Additionally, the study was designed to gather information about potential impacts on students, school personnel, and school climate (more information about this evaluation is discussed in the section that follows).

Interim findings from this study were reported to the Foundation, after which the Foundation “decided it wanted to move ahead and do a randomized control trial.” The R&E program officer worked closely with the implementation evaluator “to help develop the RFP [request for proposals] and the outcome measures that we were going to be looking at.” An

interesting feature of the RFP was a requirement that the selected evaluators would work in partnership with the research center involved in the implementation evaluation. According to the implementation evaluator this was “because we were so familiar with the program already, [the Foundation] wanted to keep us in there.” Additionally, the implementation evaluator was “a part of the selection process” to determine which evaluation firm would conduct the impact study. This process included inviting particular evaluation firms to submit formal proposals, and then conducting interviews with them over the phone.

Four firms were asked to submit proposals to address two factors of the Fun and Games program. The first was an expanded examination of the program implementation across sites, and the second was an investigation of program impacts. Two firms submitted proposals, and one firm was selected to conduct this evaluation. To assess implementation, the impact evaluator developed and utilized measures including interview protocols for program staff and observation protocols. The impact assessment was a randomized control trial with such measures as teacher and student surveys, physical activity trackers, and school records. In his/her proposal, the implementation evaluator also submitted a list of anticipated deliverables for the project, including policy briefs and final reports to the Foundation, study participants, and publicly available data repositories.

At this time, the project is still ongoing and continues to expand. Information from evaluations is shared via the web, and several journal articles are being prepared for publication.

Case 2: Challenging case. The Foundation again assisted in the selection of this second case. During initial interviews at the Foundation, it became clear that this program and the evaluation of it are perceived to be a disappointment. Similar to the success case, this program was developed in a pilot phase that was subject to an implementation evaluation. Following the pilot phase, the program was scaled up to a demonstration phase that included an impact evaluation.

Table 4 shows the data sources consulted to describe this case. There were fewer data sources available to describe this case. This is due partly to the time that has passed since the program completion, and partly because of a lack of response in some instances. It is important to note that the national program director was invited to participate in an interview but never responded to the interview requests. Instead, the program director's perspective is represented through a report that s/he filed with the Foundation, and through interviews with some of the Foundation personnel that worked with him/her throughout this project.

Table 4
Data Sources for the Challenging Case

Documents	
	Program Call for Proposals
	Implementation Evaluation Design
	Implementation Evaluation Final Report
	Demonstration Site reports
	Grantmaking Lessons (internal)
	Program Office Final Report
	Program Results Report
	Impact Evaluation Final Report
	Peer-reviewed journal article discussing impact
	Annual Reports
Interviews	
	Foundation Research & Evaluation Program Officer
	Foundation Senior Vice President
	Lead Impact Evaluator
	Foundation Program Officer
	Foundation Research & Evaluation Advisor

This section includes a brief overall summary of the history and background of this program, followed by an overview of the evaluation activities associated with it. Finally, the processes of both evaluations are presented in-depth.

Program history and description. The Drug Free Families program sought to prevent substance abuse by strengthening at-risk communities and reducing risk factors. The program was delivered through early childhood centers to target high-risk families with low-income preschool age children. The goal was to reduce young children's vulnerability to substance abuse through strengthening their families and neighborhood environments. As described by a senior program

officer, the program used, “early care and education centers in various high risk communities as a change agent for the community environment in order to reduce some of the factors that are in play within the community that put the children at risk.”

This theory of action tied into the Foundation’s strategic objectives at the time surrounding substance abuse, and its prior support of programs that focused on investing in environmental factors as a means of preventing substance abuse. The national program director of Drug Free Families had a strong and significant background in working on issues related to children and poverty and substance abuse. S/he was the Foundation’s first choice to serve as program director, and played a significant role in shaping the program. The national program director was described by Foundation personnel as passionate and hard-working with deep commitment to, and belief in, the Drug Free Families program.

Due to the fact that the goals of the program were community-wide, the targeted recipients extended beyond those families enrolled in the child care centers to other community members with un-enrolled children. Funded sites built “partnerships with local service organizations, schools, police, and other government agencies, which would work together to implement strategies based on nationally recognized best and promising practices, to strengthen families and the community.” The focus on community-level factors was motivated by several factors. First, during the pilot phase, the belief was that the partners would provide more expertise than the childcare centers about substance-abuse prevention, and would thus be able to strengthen the intervention. Second, building community partnerships was anticipated to increase the chance that program services would reach children and families even after funding concluded.

Primary program activities were divided among community, family, and partnership focuses. Families received case management (including assessment), peer mentoring (individual and in support groups), and counseling. Community efforts included advocacy, forums, and community

action projects. Partnership-focused strategies involved neighborhood revitalization and developing agreements with substance abuse organizations to treat referrals from the program.

The program began in 1992 as a pilot in five geographically diverse sites. The national program director was heavily involved in selection of sites through an open competition managed by the Foundation. The demonstration and evaluation phase began in 2001 as the program was expanded to 15 sites. The program concluded in 2005.

The work of the Drug Free Families program was overseen by two advisory committees. The first, populated with experts in related substantive fields, provided general guidance to the program and assisted inside selection. The second was an Evaluation Advisory Panel, commissioned in 2000 to guide the research and evaluation of the demonstration phase of the program. According to a program officer, the members of the advisory committees “were picked by a combination of the Foundation and the program office, and they were picked to be people who were knowledgeable and influential ... [The advisory committees] were stalwart leaders” in the area of substance abuse prevention.

Both the pilot phase and demonstration phase were the subject of evaluations conducted by different evaluation groups. As with the Fun and Games program, the pilot phase evaluation focused on implementation, and the demonstration evaluation focused both on implementation and impact. Each of the evaluation efforts is described in greater detail below.

Pilot and implementation study phase. The same firm that conducted the impact evaluation for the Fun and Games program was contracted to conduct the implementation study. The evaluation “was designed to document the model development and implementation process and to derive lessons that would help shape a full-scale demonstration of the Drug Free Families approach.” The evaluation carefully detailed community and family risk and protective factors, program context, site characteristics, program strategies, and perceived community and family

changes. The report went on to offer suggestions for ways to enhance the success of a larger roll-out by isolating particular strategies that seemed to be most effective.

Overall, the implementation study suggested that there were some promising aspects of the Drug Free Families program, but that it was implemented inconsistently across sites despite the initial strength of the selected locations. During interviews, several people suggested that the program was still struggling to define its activities and develop a concrete model. Part of the work of the evaluator during this phase was to “look at the work that had been done in the pilot phase and kind of characterize what it was. [The evaluator] had to help operationalize it; his/her role was to operationalize pilot phase results to a large part, to categorize them according to these general types of interventions.”

In hindsight, the Foundation staff agreed that the program was not ready for a large-scale roll out or an impact study. The positive findings from the implementation study centered around participant’s perceptions of the program’s relationship with desired impacts and the idea that incorporating this model of substance abuse prevention into early child care centers was feasible. At this point, there was very limited evidence to suggest the program was efficacious in reaching the desired impacts. The evaluation report documented satisfaction and perceptions from parents, teachers, and community partners. Despite the lack of information available to draw conclusions about the impact of the program, the Foundation staff and program director felt that the findings gave enough of an indication that the program had potential for strengthening communities.

An additional factor weighing on the decision to scale up was interest expressed by the federal government in investing in similar types of programs. The feeling among the program director and Foundation personnel was that there was an important political moment rife with potential for a significant federal investment. If positive impacts could be demonstrated during the

evaluation of the demonstration, perhaps these efforts could play an important role in shaping federal policy.

Looking back on this, Foundation staff realize that political will and timing were insufficient to assist the program in realizing its potential. One Foundation staff member explained, “But we didn't have the interventions nailed.” Furthermore, the evaluator’s recommendations were “weak.” However, “[The national program director] was an advocate, and s/he thought we had this one chance with the federal government to get appropriations now before it goes into kind of the cookie cutter mold.” The staff member went on to further say, “I would never develop an implementation program around these feeble findings ... [the Foundation] had four models that had no effectiveness, and we were going to model a whole program on replicating those models and fourteen communities with fourteen comparison sites.” Despite some reservations, the Foundation went on with the demonstration phase and attendant evaluation.

Demonstration and impact study. In 2000, an open call for proposals was issued by the Foundation to select the demonstration sites. Initially, the Foundation presented four successful pilot models (identified through the implementation study). However, early on in the application process, the Foundation changed these guidelines to be more explicit about the approach and subsequent activities that should be adopted by each site. This change was made at the urging of the Evaluation Advisory Panel because the members believed that having a more uniform intervention in place across sites would allow for a more rigorous investigation into the program’s impacts.

The Evaluation Advisory Panel was commissioned in 2000 to advise and oversee both the program and evaluation designs. The evaluator from the implementation evaluation was also involved in discussions, offering recommendations from the implementation study to inform both the demonstration program and evaluation. Proposals for the impact evaluation were solicited in 2001. The Foundation received three proposals. These were reviewed by the national program

director, the Foundation research and evaluation staff, and the chair of the Evaluation Advisory Panel. Ultimately, a proposal from a prominent research university was selected “as the strongest and most appropriate proposal.”

Despite the change in applications that was designed to promote more implementation uniformity across sites, the transition from the pilot to the demonstration phase proved challenging. Ultimately, program implementation varied widely from site to site. According to one report, “The sites sought, and were permitted to adapt the model to local conditions. The program therefore lacked a uniform appearance from site to site.” A request for proposals for evaluators went out to select firms following the selection of demonstration sites through a competitive open call process, “so the evaluators knew more or less (and mostly less) what was going to be evaluated, that’s when the evaluation was formed; it was around these sites.”

During the demonstration phase, the program was met with numerous challenges. The national political context surrounding the childcare centers changed drastically (including a shift in focus and a significant decrease in funding), and there was a lack of experience with and capacity for developing large-scale programs among site staff. As these challenges played out, the staff at each of the program sites needed to shift their attention from Drug Free Families implementation in order to address the other changing demands.

Fourteen of the fifteen sites participated in the impact evaluation. These were matched to 14 comparison sites to facilitate a quasi-experimental design. The evaluation measured both process and outcome indicators and documented several challenges. In addition to the inconsistent program implementation mentioned above, the challenges were overarched by the fundamental problem that the desired program effects were very distal from the intervention and incredibly complicated to measure. Ultimately, the program activities were not well-aligned to the stated goals.

These measurement challenges were further complicated by the ambitious scope of the program and the multiple and complex contextual factors that influenced implementation and the potential for impact. These challenges spawned multiple disagreements and tensions between the program personnel and the evaluators. The program was concerned that the evaluation failed to capture the “program’s real-world impacts” because it was a “pure academic research model.” The evaluator and the Evaluation Advisory Panel, however, felt that the research design was rigorous and appropriate, but that the inconsistencies across implementation “made it difficult for the evaluators to evaluate the conceptual model uniformly.”

The evaluation suggested that even though implementation varied widely, most of the programs were able to implement various strategies with some degree of success, and that the program did contribute to establishing community partnerships between target organizations and participating childcare centers. However, on measures of impact, there was little or no evidence that treatment sites were more able than comparison sites to bring about changes to families.

Ultimately, the findings were inconclusive. One Foundation staff member summarized this idea, saying, “We can’t say that family and community strengthening didn’t work. All we can say is that in this particular demonstration we didn’t show clear effects, and few sites fully implemented the model. This doesn’t nullify the idea.” The program was unable to demonstrate effectiveness to the government such that they could influence appropriations. The program closed in 2005.

Case 3: Complex case. This case is one of the Foundation’s largest programs and evaluations currently in operation. Unlike in the other two cases, there was one evaluation team commissioned for the duration of the program. Implementation and impact studies were integrated, and the evaluation had to continually evolve to match the changing and expanding program. Both the program and evaluation are ongoing. Because of its size and long tenure with the Foundation,

many different Foundation staff members have been involved with the project. Table 5 shows the data sources consulted in developing this case.

Table 5
Data Sources for Complex Case

Documents	
	Program Call for Proposals (2006, 2008, 2009, 2010)
	Evaluation Request for Proposals
	Evaluation Proposals (2008, 2009, 2013)
	Evaluation Contract Summary
	Letter of Agreement between Foundation and Evaluator
	Evaluation Timeline
	Evaluation Plan
	Community Snapshot Reports
	Various "Lessons Learned" Reports from Evaluator
	Various Peer-Reviewed Journal Publications
	Special Issue of a Journal Devoted to Evaluation Findings
Interviews	
	Foundation Program Officer
	Foundation Program Officer
	Foundation Research & Evaluation Program Officer
	Foundation Research & Evaluation Program Officer
	Foundation Research & Evaluation Program Officer
	Foundation Senior Vice President
	Lead Evaluator

Program history and description. In an effort to build more coordinated, collaborative, and effective healthcare systems in communities, the Holistic Health program “builds multi-stakeholder partnerships—among insurers, providers, purchasers, and consumers—for the purpose of improving patient health and health care.” The program began in 2006, and will continue in its current form until 2015. According to Foundation leadership, the program efforts represented an effort to take, “a place-based approach towards improving quality with the rationale that healthcare is local, and healthcare problems can’t be addressed by having a program that impacts one hospital in one place and a floor of another hospital in another place. What [the Foundation] needed to do was to move across the market and have an impact that would move the quality needle in a measurable way across the entire market.”

The program initially began in four markets. Locations were selected based on the Foundation's initial scans of communities according to pre-set criteria. The Foundation was interested in building on current systems and gathering information that would inform a regional approach. After less than a year had passed, the Foundation decided to scale-up and asked communities to submit proposals for the expansion. The evaluator of the program explained, "initially the program was meant to start off as a pilot and then maybe eventually moved to scale. [The Foundation] chose four initial grantees that were supposed to be pilots, but didn't really follow that process. [The Foundation] basically said we are going to scale this up before there was ever an opportunity to even learn from the pilots."

Throughout the years the program has been in operation, there has been some fluctuation in the number of funded communities. Eventually, it settled at 16, all still in operation. This strategy was explained by a Foundation employee, "I think the first phase was sort of trying to decide what was important so that we could do the selection, so we could look at characteristics and the pilots the communities (and identify) what we think were important characteristics that helped frame the call for proposals for the next phase."

The specific program activities span five areas: 1) measuring and publicly reporting healthcare quality; 2) engaging consumers as partners in their own care; 3) spreading effective quality improvement strategies; 4) ensuring healthcare equity; and 5) integrating these activities with payment reform. Ultimately, the goal of the program is to "improve quality of healthcare and healthcare delivery as well as value for healthcare services."

The evaluation team was involved in various ways throughout the process. One Foundation staff member explained, "we wanted to get the evaluators involved as we were deciding the program so that we could think through the metrics. So the evaluation really began before the group first

group were selected and was continuous throughout, giving feedback as we were making programmatic decisions.”

Evaluation. The primary evaluator is a faculty member at large, prestigious state university, directly identified by the Foundation and invited to submit a proposal based on his/her previous research and evaluation work in the same content area. This happened before any of the individual sites were funded, so the ideas for how the program would operate were still nascent. The program officer responsible for the contract recalls, “I remember at the time it was very hard to explain what the program was because it was constantly changing. So the RFP was particularly challenging to explain what exactly [the evaluators] were going to be evaluating and what kind of role they were going to play.” Because the program was so large, the evaluator assembled a team of experts to work on it. The evaluator estimates that as many as 40 people have worked in some capacity on the evaluation.

According to one research and evaluation officer, the goal of the evaluation is “to find out the extent to which communities publicly reported their quality measures. And at a later point in time, to look and see whether this has an effect on consumer decision-making and on quality improvement in healthcare institutions and medical decision care making in general.”

Given that the Foundation decided to scale up almost immediately after the program began, the evaluation team had to adapt quickly. The evaluator was praised by many at the Foundation for his/her flexibility and responsiveness to the changing program and climate. Program and evaluation activities are ongoing, so there are not yet final findings to share. However, the evaluation team has been prolific in sharing portions of the data and stories of the evaluation process. The evaluator estimates that around 60 peer-reviewed articles have been published in addition to multiple issue briefs.

Analytic Procedures

Analysis of the data collected in this study occurred in iterative phases both during and after data collection activities. This section provides information about the deductive and inductive data analysis procedures used to answer the study's research questions. This study relied heavily on a coding framework developed in the early phases of the research. The process for its creation is described first in this section. This is followed by the specific analysis procedures applied to each of the different data sources: records, interviews, and case studies.

Coding Framework

Three previously existing frameworks informed the coding scheme. Two of these were developed specifically to describe different facets of evaluation policy (American Evaluation Association, 2007; Trochim, 2009), and the other was used in a study that created logic models describing contrasting theories of evaluation (Hansen, Alkin, & Wallace, 2013). The last framework was used to inform the overall structure and organization of the present coding scheme, and the first two were called upon to fill in the critical content of evaluation policies. The coding framework was drafted, then piloted with several raters during a process devised to refine it. The following paragraphs describe this process in greater detail.

The Hansen et al. (2013) framework was developed for the purpose of using logic models to depict various evaluation theories or approaches. As defined in the first chapter, a logic model (most often employed to facilitate understanding of a particular program or intervention) depicts the components of a program (inputs, activities, and outputs) and the logical connections that tie those components to desired outcomes. Even though this application does not directly match the goals of the present study, the broad categories—contexts/inputs, activities, outcomes, and assumptions—are also applicable to this investigation of evaluation policy. For the purposes of this study, the researcher is conceptualizing evaluation policy in these terms commonly associated with logic

models. That is to say that the broad definition of evaluation policy guiding this research includes guidelines for determining the inputs, activities, and desired outcomes of evaluation.

As further explanation, evaluation policies, like programs, are designed with particular desired outcomes in mind. These outcomes are supported by particular activities, the links between which are supported by a theory of action that makes the logical connections more explicit. Specifically, the inputs of evaluation policy include the guidelines describing evaluators (their skills, knowledge, competencies, values, etc.), evaluations (determining evaluands, timing evaluations, and defining the practice), resources (financial, human, and otherwise), and stakeholders (their identity and the rules that determine their inclusion). Activities describe the evaluation policies related to planning and preparing the evaluation (understanding the program, designing the evaluation, and designating roles, etc.), conducting the evaluation (data collection, data analysis, communication, and developing conclusions, etc.), and reporting (content, format, and audience). Evaluation policy outputs include the goals of the evaluation policy, findings, and use of evaluation results. Finally, assumptions speak to the beliefs about program, evaluation, epistemology, stakeholders, methods, etc. that underlie evaluation policy.

Two other sources were consulted in the development of the framework. The first was the American Evaluation Association Evaluation Policy Task Force. This body was commissioned in 2007 and charged with assisting the American Evaluation Association with identification of, and influence over, evaluation policies that influence evaluation practice (American Evaluation Association, 2007). This charge included developing a definition of evaluation policy that outlined the broad areas that might be included in evaluation policies. These are: evaluation definition, requirements of evaluation, evaluation methods, human resources regarding evaluation, evaluation budgets, evaluation implementation, and evaluation ethics.

These broad categories of evaluation policy were introduced when appropriate into the coding framework. For example, evaluation definition was not present in the initial Hansen framework, but seemed appropriate for this study. In these cases, the codes from the EPTF were added to the framework. The same process was introduced with the evaluation policy framework presented by Trochim (2009). As a reminder, the categories he presented in his evaluation policy wheel are: evaluation goals, evaluation participation, evaluation capacity building, evaluation management, evaluation roles, evaluation process and methods, evaluation use, and meta-evaluation policies. These categories were introduced into the framework where appropriate. The mapping of the additions to the evaluation theory framework is presented in Figure 2.

Figure 2
Evaluation Policy Domains Represented by Evaluation Policy Coding Framework

Framework:	Hansen et al.	Trochim	EPTF
Evaluation Context	Evaluator Program Stakeholders General	Evaluation participation	Requirements of evaluation Human resources Evaluation budgets Evaluation Implementation Evaluation definition
Evaluation Activities	Planning/Preparation Conducting evaluation Reporting Outputs	Evaluation process and methods Evaluation capacity building Evaluation management Evaluation roles	Evaluation methods Evaluation Ethics
Evaluation Outcomes	Characteristics of the evaluation Organizational capacity Use External factors	Evaluation Use Meta-Evaluation Evaluation Goals	
Evaluation Assumptions	Program Epistemology Evaluation Stakeholders Other		

Several areas of both consistency and inconsistency are evident across these three frameworks. Additional work was required to combine, refine, and further flesh out these coding frameworks. This process was largely derived from the researcher's experiential knowledge working as a program evaluator. Scholars and practitioners of evaluation were consulted to further refine the final version of the framework, which will henceforth be called the Evaluation Policy Coding Framework, or EPCF.

The next step in the process was piloting the EPCF with real data from the study. With the assistance of four raters, the EPCF was utilized in a preliminary examination of the Foundation's

policy records. The raters were assigned different documents to code. Decisions were made about which codes to delete and which to add at in-person meetings between the raters and the researcher over a six-week period in the spring of 2013. The final framework, presented in Figure 3 (with definitions in Appendix J), served as the basis for the rest of the analysis in this study.

Figure 3
Final Evaluation Policy Coding Framework

Evaluation Policy Assumptions	Evaluation Activities
Program	Capacity building
Epistemology	Planning/Preparation
Stakeholders	Contracts
Evaluation	Understanding the Program
Methods	Designing the Evaluation
Results	Roles
Reporting	Conducting the Evaluation
Use	Methods
Evaluation Context/Inputs	Evaluator's Role
Program	Foundation's Role
Evaluation	Program's Role
Definition	Administration
Schedule	Communication
Evaluator	Managing Relationships
Selection	Stakeholder Participation
Technical Expertise	Data Collection
Content Knowledge	Data Analysis
Interpersonal Skills	Developing Conclusions
Values	Findings
Stakeholders	Ethics
Identifying Stakeholders	Reporting
Stakeholder Inclusion Reason	Format
Resources	Contents
Time	Audience
Financial	Metaevaluation
Personnel	Evaluation Effects/Consequences
	Goals
	Findings
	Use
	External Factors

Record and Interview Analysis

This study began with analysis of the documents and videos collected to describe the evaluation policies of the Robert Wood Johnson Foundation and the initial interviews conducted in May of 2013. These materials were analyzed with the assistance of Dedoose, software designed to support the management and analysis of qualitative data. Three phases of coding were employed to analyze the data. Each of the techniques is described below.

Attribute coding. The process began with attribute coding (Saldaña, 2009) to note basic descriptive information about each record. In the case of the records, notations included author, year of creation, intended audience, and overall topic. For the interviews, information about department, title, and length of service at the Foundation was noted. This was an essential step in the analysis process because it provides information and contexts that support further analysis.

Provisional coding. The first substantive phase of coding was provisional coding, which began with application of codes from the EPCF. Provisional coding initiates with a predetermined ‘start list’ of codes before data collection (Miles & Huberman, 1994). The process for generating this list was detailed in the description of the development of EPCF above. An important note about provisional coding is that “as qualitative data are collected, coded, and analyzed, Provisional Codes can be revised, modified, deleted, or expanded to include new codes” (Saldaña, 2009, p. 144) The records were loaded into Dedoose and information about their attributes was indicated. Then, excerpts were selected and tagged with provisional codes.

Elaborative coding. Once provisional coding was complete, the second stage coding technique, elaborative coding, was employed. This process of analyzing the data contextually facilitated further development of the theory. Saldaña (2009) indicates that this style of coding is appropriate for qualitative studies that work to build on previous research and investigations in an effort to “support, strengthen, modify, or disconfirm the findings from previous research” (p. 229).

These two coding stages were instructive in finalizing the EPCF, which is presented in the findings chapter. When the applications of codes to excerpts were finalized for this phase of the research, data from the case studies was analyzed.

Case Study Analysis

The general analytic strategy used for the case studies relied on theoretical propositions that guided the research questions which led to this case study (Yin, 2009). The first step in the analysis was descriptive in order to present a comprehensive account of the relevant features in these cases. This was followed by theoretical analysis, which applied the general coding categories from the Evaluation Policy Coding Framework to excerpts from both the documents and interviews. Establishing these categories allowed for a cross-case synthesis according to different evaluation policy domains. This facilitates a robust discussion of the implementation of various Foundation evaluation policies.

CHAPTER 4

DESCRIBING THE FOUNDATION EVALUATION POLICIES

Introduction

In this chapter, the findings from Phase I of this study are presented. First, a brief overview of the data collection methods is reviewed, followed by a discussion of the analysis procedures. Finally, the findings from the document analysis and key staff interviews are presented. Specifically, the findings in this chapter address the following research questions:

- What are the evaluation policies of the Robert Wood Johnson Foundation?
 - What do the policies communicate about stakeholder involvement, evaluation methods, evaluation goals, etc.?
 - Through what means are the evaluation policies communicated to those responsible for carrying them out?
 - How are the evaluation policies developed and determined?

Answering these questions is far from simple. The Foundation does not have one document to consult that unequivocally outlines its evaluation policies; it does not approach evaluation policies in this direct sort of way. Rather, policies emerge through careful consultation of a variety of sources—the history and culture of the organization, published and internal records that describe practices, and in the minds of the employees. This chapter presents a consultation of each of these sources in an attempt to draw out the Foundation policies that guide the research and evaluation work that takes place. Together, the records and interviews afford a variety of perspectives and insights into the evaluation policies of the Robert Wood Johnson Foundation. What follows is analysis of this data to address the first four research questions in this study.

The Evaluation Policies of the Robert Wood Johnson Foundation

As a reminder, this study is guided by an extremely broad definition of policy: an evaluation policy is “any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation” (Trochim, 2009, p. 18). Another important aspect of Trochim’s definition is that policies can manifest in a variety of formal and informal ways—in order to be a policy, it does not have to be explicitly presented as such. This is largely the case regarding the Foundation’s evaluation policies. A sense of the policies can be gained by consulting a variety of sources and people, in order to draw inferences. An overall trend that emerges from consulting the data is that there are two different categories of policies: those that are goal statements and those that are operationalized activities. Both satisfy Trochim’s (2009) definition, but play out in very different ways in practice.

Evaluation policy goal statements broadly refer to what evaluations are supposed to accomplish. For example, a phrase repeated commonly in the records is that evaluations are supposed to “build an evidence base around a strategy.” Statements like these describe the Foundation policies’ ultimate objectives. In theory, the specific evaluation activities should be undertaken in ways that contribute to this goal. The specific activities, then, are the operationalized policies. Some examples (that will be discussed further in this chapter) are about stakeholder involvement and report writing. In both cases the Foundation has produced explicit guides that document what evaluators should do in particular situations. Sometimes operationalizations connect to the goals in explicit ways; however, this isn’t always the case.

This distinction between policy goal statements and policy operationalizations might be further clarified by an example of a well-known education public policy—The No Child Left Behind Act (No Child Left Behind [NCLB], 2002). Like many other policies of its kind, NCLB is described in terms of its broad goals—that by 2014 the achievement gap between minority students and their

White classmates would be eliminated, and all children would score proficient or above on standardized testing measures. The operationalization of this policy—the specific means by which these goals are supposed to be achieved is left up to the states and school districts. This process is mirrored somewhat at the Foundation. While there is tremendous alignment between the records and among the interviewed staff about the goals of the policies, their operationalization is left up to each program officer resulting in variable understanding, adherence, and enactment of any existing operational policies.

I can offer another analogy to help clarify the distinction between goal statements and operationalizations: the relationship between them should mirror relationships between activities and outcomes of programs. A program's activities are shaped by the outcomes that the program is trying to accomplish. Ultimately, there should be a clearly defined and logical connection between the activities encouraged by the policies and the desired outcomes. As the data show, there is variability as this plays out in practice in the Foundation's evaluation policies. A more in-depth exploration of these different types of policies helps to answer the next research question, "What do the policies communicate about various aspects of evaluation practice?"

What the Evaluation Policies Communicate about Various Aspects of Evaluation Practice

One of the first steps in this study was to develop a coding framework to assist with this portion of the analysis (the process entailed in this is described in the methods chapter). A deductive approach to analyzing the data was purposely chosen in order to link this empirical investigation to the theoretical work that has previously been done. Deriving a codebook in advance also affords an opportunity to discuss aspects of the Foundation's evaluation policies that are "missing."

While the most granular level of coding still holds up, during the analysis, the organizational framework for the codes proved not to be a great match to the ways in which the Foundation's policies are structured. Rather, several aspects of evaluation processes emerge. These aspects

include: (1) evaluation goals; (2) evaluation assumptions; (3) determining what to evaluate and when evaluations should be conducted; (4) selecting and contracting evaluators; (5) designing and conducting the evaluation; (6) administration; (7) stakeholder involvement in evaluations; and (8) reporting evaluation findings. Table 6 presents the reorganized coding framework based on these domains in addition to the number of codes that appear in the document analysis.

Table 6
Codes Nested in Process Categories

Coding Categories		<i>n</i>	%
Goals			
	Outcomes	12	
	Use	24	
	Meta Evaluation		
Subtotal:		36	8%
Assumptions			
	Use	6	
	Methods		
	Reporting		
	Evaluation	8	
	Epistemology	5	
	Program	1	
	Stakeholders	4	
	Results	3	
	External Factors	5	
	Other	2	
Subtotal:		34	8%
Deciding what and when to evaluate			
	Identification of programs to evaluate	1	
	Schedule of evaluations	13	
	Definition of evaluation	17	
Subtotal:		31	7%
Selecting and contracting evaluators			
	Contracts and Proposals	7	
	Time	4	
	Personnel	1	
	Financial Resources	15	
	Evaluator values		
	Evaluator selection	8	
	Evaluator technical expertise		
	Evaluator interpersonal skills		
	Evaluator content knowledge		
Subtotal:		35	8%
Designing and conducting the evaluation			
	Understanding the program	14	
	Designing the evaluation	43	
	Methods	8	
	Data collection	9	
	Data analysis	7	

	Findings	3	
	Developing conclusions	2	
	Ethics	1	
Subtotal:		87	20%
Administration			
	Communication	26	
	Managing relationships	14	
	Foundation's role	5	
	Program's role	7	
	Administration	8	
	Evaluator's Role	10	
	Roles	9	
Subtotal:		79	19%
Stakeholders			
	Stakeholder Inclusion Reason	26	
	Identifying Stakeholders	12	
	Stakeholder Participation	33	
	Capacity Building	7	
Subtotal:		78	18%
Reporting findings			
	Format	11	
	Other	2	
	Contents	11	
	Audience	13	
	Findings Characteristics	8	
Subtotal:		45	11%
Total:		425	

Note. Numbers only appear in the percentage column next to subtotals because they indicate the proportion of codes relative to the entirety of the sample at the domain, rather than category level.

The division between evaluation policy goal statements and evaluation policy operationalizations will frame the discussion of the data presented in this chapter. It begins with a description of the evaluation policy goals as they are evidenced in the records and interviews. Then, the assumptions held by the Foundation that underlie these policies are discussed. Next, the above-mentioned evaluation processes governed by policies are explored. The chapter concludes with discussions about how these evaluation policies are developed and communicated to people within and outside the Foundation.

Evaluation Policy Goals

The goals of the evaluation policies are a critical place to begin the discussion detailing evaluation policies. What the Foundation hopes to accomplish through its evaluation activities provide the basis for setting policies and rules that guide evaluation practice. Table 7 displays the various codes that exist in the record analysis. Specific descriptions of the desired qualities of the information gathered through evaluation activities and the ways in which the information should be used are detailed in these excerpts.

Table 7
Codes Pertaining to Evaluation Policy Goals

Goals			
	Use	24	67%
	Outcomes	12	33%
	Meta Evaluation	0	
Total:			36

Note. The codes in this domain represent 8% of the overall coding framework.

Of primary importance to the Foundation is the use of evaluation findings. This is evidenced by the fact that two-thirds of the codes applied in this category describe various aspects of use. The Foundation's chief interest when it comes to evaluations is that findings are produced and used to inform its grantmaking practice and the field in general. Given that it is the nation's largest Foundation dedicated to public health, it takes very seriously its responsibility to advance high-quality practice. The Foundation holds a firm belief that evaluation is a key part of that strategy. Repeated multiple times in interviews and the documents is the idea that all findings from evaluations are useful in some way, regardless of whether or not they are positive or negative. More specific information about the excerpts contained in this section is examined below.

Use of evaluation findings. The Foundation devotes a significant amount of attention to the concept of use of evaluation findings. The most common themes to emerge in this set of

excerpts are the concepts of “learning” or “informing.” Some specific quotes that support these ideas include:

We design [evaluations] to capitalize on the greatest opportunity for learning, given the nature of the program and strategy being examined.

The purpose of Foundation evaluations is to learn from a program’s policies or strategies.

The Foundation does evaluations for learning.

The concept of learning is described as the primary objective of evaluations. This seems the highest level of outcome, the most vague, and least able to be operationalized. As the Foundation moves to discuss greater levels of specificity, the talk of “learning” moves to talk of “informing.” The chief difference between the two is that informing has a particular audience in mind. Specifically, in these excerpts, the subjects of informing are the field, practice, and policy. Many of these are mentioned in conjunction with one another in the same excerpt. Informing the field and informing policy appear in four of the excerpts, and informing practice appears in five excerpts. Three of these excerpts include references to informing both policy and practice. Informing the field, however, never appears concurrently with either informing policy or informing practice.

Though “policy,” “practice,” and “the field” appear multiple times in the excerpts, the terms are never specifically defined. The terms most likely represent people in some way—policy makers are those who hold office or work in some federal capacity, and practitioners are people engaging in the work implementing and executing programs or services. The field, on the other hand, is more ambiguous. The Foundation might be using it to refer to the combined ideas of policy and practice (particularly because it is never mentioned in addition to either of those two things), but it could also include academics, researchers, the general public, consumers of similar services, etc. Some examples of references to the field include, “The goal is to learn and to inform the field,” and, “[The objective is]to inform the field of work we are trying to help build.”

The ideas surrounding use of evaluation findings are strongly echoed in the interviews conducted with Research and Evaluation staff and Foundation leadership. During the interviews, a question asking participants to describe the Foundation's policies was posed. Frequently, the initial responses were that the Foundation "doesn't have an evaluation policy." However, when prompted to explore this topic further, most of the participants (89%) made reference to some sort of learning or informing—it is very consistent with how use of evaluation findings are described within the documents.

Something that emerged in the interviews that was not apparent in the document analysis is the idea of organizational culture and how the history of the Foundation influences the attitudes towards the goals of the evaluation policies. Evaluation has been woven into the fabric of the organization since its inception. The original founders and staff members placed high value and emphasis on evaluation, and that persists culturally within the organization to date. As one Foundation leadership member explains, "I think a broad level we've always operated [on is that] we've seen ourselves as a learning organization. We've been dedicated to understanding the implications of what we've done, and using that as an additional way to guide our future steps. So from the very start, the Foundation has seen evaluation as important, for several reasons: (1) to guide our future efforts; and (2) to provide this information to the field."

The Foundation's attitude towards the use of evaluation findings is clearly informed by this culture and history. It is upon this basis that the Foundation's evaluation practices are built. The other excerpts tagged with codes in this category describe in more specific detail the characteristics of evaluation outcomes, and the Foundation actions they support.

Outcomes. Excerpts in this category also describe the desired outcomes from practicing evaluation. Although use of findings, as described above, is clearly an important outcome, the excerpts tagged with "outcomes" differ somewhat. Specifically, these excerpts describe more in

detail what the Foundation will be able to do differently as a result of engaging in evaluation. The general themes encompassed within this section include the Foundation's ability to assess the strengths and weaknesses of the strategies it employs, to understand its impact, to spread effective approaches, to ensure its own credibility, and to promote social change. Some excerpts that further explain these concepts, include:

Well, evaluations are really an important part of our national reputation for credibility. When you ask outsiders what makes us credible, research and evaluation is a big part of that.

When evaluation works well, it provides information to a wide range of audiences that can be used to make better decisions, develop greater appreciation and understanding, and gain insights for action.

Since RWJF represents a public trust, it is essential that it understands the perceived impact of its work on all of the affected parties

These provide a link between general attitudes and philosophy about use of evaluation findings to what the evaluation hopes to do with the information that it gathers.

Metaevaluation. The final subcategory in this section is metaevaluation—the evaluation of evaluation. This is an area in which the Foundation does not have any specific policies represented in the documents consulted for this section of the analysis.

Assumptions

Assumptions shed light on the underlying beliefs that explain how the goals of the evaluation policies are going to be accomplished. In the program evaluation literature, this is oftentimes referred to as a theory of action (Alkin, 2011). When considering a particular intervention, this generally represents the mechanisms by which a program is expected to accomplish its goals. The theory of action is what undergirds “If... then...” statements. For example, consider a teacher professional development program utilizing an observation-based

coaching model. The program's long-term goal is to improve student test scores. The statement would be, "*If* teachers receive individualized expert coaching, *then* student outcomes will improve." Two assumptions underlie this statement: (1) coaching teachers will lead to improved pedagogical practices; and (2) improved pedagogy relates to better student outcomes.

For the investigation at hand, assumptions explain the link between the Foundation's evaluation policy goals, and the operationalized activities. Assumptions are discussed in depth in the text that follows, but an example upfront might help clarify. As mentioned previously, the Foundation is very interested in promoting different types of use of evaluation findings. The Foundation believes that one way to enhance use is to involve program stakeholders in the process. A related theory of action statement would read, "*If* stakeholders play a role in designing the evaluation, *then* they will be more likely to use findings." Several assumptions underlie this statement. The first is that program personnel (stakeholders) are in a position to use evaluation findings. The second is that stakeholder involvement in the process will increase their buy-in, and the credibility of the eventual findings. Finally, these factors will ultimately lead to the stakeholders' desire to use the evaluation results.

Assumptions are rarely written or stated directly. More often, ascertaining assumptions relies on inference based on statements that convey the theory of action. This section begins with an exploration of the assumptions as they are conveyed through the records analyzed in the first phase of the research. This lays the ground work for the discussion of the more specific process operationalizations that follows.

Table 8 shows the number of excerpts tagged with codes that represent beliefs held about various evaluation components that inform the policy decisions. This part of the coding represents 8% of the overall framework and includes several categories. One of the categories in this domain is evaluation. While it is obvious that assumptions about evaluation will influence the evaluation

policies, beliefs about evaluation use, epistemology, stakeholders, results, program, methods, and reporting also bear importantly on evaluation activities. Each of these categories of assumptions are explored further in the text that follows.

Table 8
Codes Pertaining to Evaluation Policy Assumptions

Assumptions		
Evaluation	8	24%
Epistemology	5	15%
Use	6	18%
Stakeholders	4	12%
Results	3	9%
Program	1	3%
Methods		
Reporting		
External Factors	5	15%
Other	2	6%
Subtotal:	34	

Note. The codes in this domain represent 8% of the overall coding framework.

Evaluation. Evaluation is consistently described in the excerpts here as an activity that supports learning. The assumption being that evaluation, when properly conducted, assists in promoting new understandings about programs. A related assumption indicated in these excerpts is that the new learning generated through evaluation activities will promote better decision-making, “develop greater appreciation and understanding, and gain insights for action.” Another assumption indicated among these excerpts is that evaluation is a challenging activity and that many obstacles are likely to be encountered as the work unfolds.

These assumptions describe fundamental beliefs that the Foundation holds about the practice of evaluation. These beliefs primarily inform evaluation design aspects—the Foundation’s assumption about what evaluation will accomplish influences what information is gathered through evaluation practice. This is more specifically informed through the Foundation’s assumptions about epistemology.

Epistemology. The excerpts tagged with this code reflect the Foundation’s beliefs about how knowledge is acquired. This has an important relationship with evaluation because it relates to potential judgments about credibility issued to evaluations based on how and what information the evaluation acquires. In several instances, these excerpts describe the Foundation’s views on research⁶. Specifically, the Foundation believes that, “Research provides our society with new road maps and paths to address our most pressing problems.” Furthermore, the process of research is explained as, “a process of narrowing broad conceptual inquiries into concrete questions that can be explored, if not answered, through gathering and analyzing data.”

Another theme that emerges prominently among these excerpts is the assumption held by the Foundation that research and evaluation are very challenging activities. This appears three times out of the five excerpts. Specifically, the assumption is that research frequently veers off course; however, “Detours may yield unexpected but important insights and lessons.” Finally, the Foundation notes that, “few discoveries take place on schedule—innovation by its nature pushes researchers into unknown territory.” These are important assumptions relating to evaluation activities because they convey an expectation that activities will not go according to plan. In fact, deviations from the plan may yield more valuable information. This stance should convey to evaluators the importance of anticipating problems and developing contingency plans—a policy that is made explicit in the evaluation activities section.

Taken together, assumptions about evaluation and epistemology provide guidelines that instruct evaluators about the designs they should employ in the service of evaluating Foundation-funded programs. Some of these guidelines are more explicitly articulated as design issues are operationalized.

⁶ In this section, research and evaluation are treated synonymously. While they are different activities, the fundamental epistemological beliefs that underlie them represent a general inquiry philosophy.

Use and stakeholders. As indicated above, the concepts of stakeholder involvement and use of evaluation findings are interrelated. The Foundation sees a clear tie between the two ideas—all of the excerpts tagged with the use code are also tagged with a stakeholder code. The key basic assumption here is that when stakeholders are involved in the evaluation process (particularly in the development of evaluation questions), findings are more likely to be used. This assumption undergirds many of the operational policies that play out regarding stakeholders. Essentially, the belief that stakeholder involvement enhances use of evaluation findings informs the selection of stakeholders and the ways in which they are involved in evaluation practice.

There is tremendous emphasis on both of these concepts (stakeholder involvement and use of evaluation findings) separately, which buttresses the importance of the connection between these two themes. These ideas are elaborated in a later section. In the above codes, there is only one instance when the stakeholder code was applied without a use code. This excerpt communicated an assumption about stakeholders and why they might be motivated to participate, namely because, “they are interested in the issue and would like to have their voices heard.”

Results. The assumptions relating to findings or results seem to serve as a mediating concept between the above pair of stakeholders and use. Many of the same themes are echoed here, but it further adds explanation for how stakeholder involvement contributes to use of evaluation findings. Specifically, when stakeholders are involved in the evaluation design, an assumption is that the results will be more meaningful, and therefore, suggested changes will be implemented with less resistance. Another assumption evident in these excerpts is that results will be perceived as more credible when the process for obtaining them is made transparent.

Program. One excerpt was tagged with a program assumption code. In it, the Foundation explains fundamental beliefs about how programs come to be, indicating, “All programs and initiatives are the result of political decisions—a belief that resources should be allocated to solving a

particular problem, or seizing a unique opportunity.” This assumption has an important relationship to evaluation activities because it represents acknowledgement of the particular milieu surrounding the interventions that are the targets of evaluations. If evaluators take this position, they will likely be attentive to this political context in their designs.

Methods and reporting. In the document analysis, there are no instances in which assumptions about methods or reporting are identified. However, it is likely unreasonable to conclude that this means there are no assumptions widely held by the Foundation that underlie these aspects of evaluation practice. For example, despite no direct statements indicating assumptions related to reporting, some statements in some of the other sections of this coding framework seem to imply an assumption. As illustration, given the emphasis placed on reports being understandable, it might be reasonable to infer that the Foundation holds an underlying belief that when reports are more understandable, they are more likely to be used. This inference also relies on other statements included in this coding framework that suggest the Foundation’s commitment to evaluation use. This logic of this interpretation may be somewhat tenuous, but it does call into question what unstated assumptions may be inferred from careful reading of the other policies indicated here.

External factors. Aside from the assumptions that underlie the evaluation policies, the Foundation also notes that external factors may contribute to evaluation practice. These factors surround and influence the evaluation activities, but are things over which the Foundation has no control. Fundamentally, evaluations always exist in a political context and the Foundation reminds its evaluators of the political nature of all evaluation activities. Aside from this general statement, the records offer a host of potential factors that might derail an evaluation, and suggestions to evaluators for attempting to anticipate some of these problems and building contingency plans into their evaluation design. This attitude was also prevalent in interviews as program officers frequently

mentioned how common it is that plans must change in response to some factor that did not proceed as anticipated.

Other. The two excerpts that conveyed assumptions, but did not fit in the above taxonomy (and are thus coded “other”) detail beliefs held by the Foundation about philanthropy and its relationship to society. In the first excerpt, the Foundation explains the critical link between evaluating philanthropy’s impact and maximizing its contribution to social progress. The second excerpt conveys beliefs about the relationship between stakeholders and philanthropy, suggesting that the latter can become isolated from the former. This belief then necessitates a policy that systematically seeks out stakeholder perspectives in Foundation activities.

As seen in the previous section, the goals of the Foundation’s evaluation policies are primarily to learn and build evidence around strategies aimed at improving conditions for the public’s health. The assumptions articulated here outline a philosophy that the Foundation has about the relationship between evaluation and its philanthropic efforts. First, the Foundation sees its responsibility as playing a role in alleviating some of society’s challenges. RWJF approaches this by funding programs designed to solve particular problems. Then, information about these programs is gathered and shared to inform practitioners and policy makers about effective strategies. This is accomplished through evaluations designed to generate new knowledge. Evaluations are strongest when they involve people who are most directly involved in the program.

These assumptions, or underlying beliefs, tie the goals of the Foundation evaluation policies to the way in which the activities are operationalized. These specific operationalizations are discussed next according to a loose chronology about the order in which evaluation activities take place. Specifically, the discussion begins with an examination of how the Foundation decides what and when to evaluate. This is followed by policies governing the selection and contracting of evaluators, and a discussion of designing and conducting evaluations. Next, the administration of

evaluations, and guidelines for involving stakeholders are presented. Finally, policies about reporting evaluation findings are discussed.

Deciding What and When to Evaluate

Seven percent of the codes from documents in the overall framework relate to processes involved in initiating evaluations. These are subdivided into more specifically defining evaluation, identifying which programs to evaluate, and scheduling the evaluations (Table 9). Each of these is explored further below.

Table 9
Codes Pertaining to Initiating Evaluations

Deciding what and when to evaluate		
Definition of evaluation	17	55%
Schedule of evaluations	13	42%
Identification of programs to evaluate	1	3%
Subtotal:	31	

Note. The codes in this domain represent 7% of the overall coding framework.

Defining evaluation. The Foundation takes care in defining evaluation across several domains. First, evaluation is distinguished from monitoring and program-level reporting. Instead, evaluation is described as a range of activities that serve to examine the processes and impacts of programs. Evaluations also vary in scale and scope. Some are small-scale single-program examinations designed for formative purposes, while others are large-scale multi-site endeavors with the aim of contributing information that informs Foundation objectives and the relevant academic fields at large.

Scheduling evaluations. The excerpts identified as relating to the scheduling of evaluations describe when evaluations should take place. These are generally outlined in terms of the stage of development and general maturity of the program that is the target of the evaluation. The Foundation is strategic about what types of evaluations should be scheduled and when. In the early stages of program development, evaluations of impact are deemed inappropriate. Rather, the

Foundation indicates that evaluability assessments and/or formative evaluations are the best approaches under these circumstances. Once a program is mature, the Foundation designs evaluations to assess program outcomes. Finally, when multiple mature programs with similar objectives have been operating for a while, the Foundation may conduct a cross-initiative evaluation with the goal of contributing knowledge that enhances understanding of particular ideas or approaches to addressing certain problems.

This idea of matching the evaluation to the stage of program development presents an example of strong alignment between what is present in the evaluation policy records and what was said throughout the interviews. When talking about designing various evaluations, program officers frequently mentioned taking the stage of program development into account. Furthermore, the interviews offered greater clarification about what programs specifically are the objects of evaluation. Several participants note that any program with an operating budget over \$400,000 will receive an evaluation, as will programs with more direct ties to the strategic objectives of the Foundation. This relates to the goal of “trying to build an evidence base around a strategy.” The interview participants confirmed that this rule is not written—it is more of a common understanding among employees.

Another issue that emerges in the interviews is the idea that the evaluation activities should begin in concert with the start of program operations. According to the Foundation staff, this has been a somewhat recent shift in the approach to evaluation. Even though this particular timing of evaluations was referred to consistently across interviews, it is not written in any of the policy records.

Program identification. Finally, there is one code that describes program inputs as they relate to evaluation policy. The excerpt is a description of various program qualities that might impede the effective implementation of an evaluation. Some of the issues raised in this excerpt include lack of program personnel investment and compliance, competing interests, and irregular

program implementation. These ideas are captured elsewhere in this domain, particularly in the excerpts that describe evaluation scheduling. The Foundation documents are consistent in describing that the stage of maturity of program development should be a key indicator for what type of evaluation should be conducted.

One theme that emerged in the interviews related to the idea of the program context is that the Foundation makes every effort to inform the program in careful detail about the expectation that the program must fully participate in research and evaluation activities in order to secure Foundation funding. Efforts are made to be as specific as possible, particularly about the data requirements for the program (if they will be asked to supply data or participate in survey collection, for example). In program contracts (separate from evaluation contracts), there is a common paragraph describing the need for program complicity in evaluation efforts.

Once it is determined that a program should be evaluated, an evaluator must be selected and contracted to work on the project. The policies governing this process are explored next.

Selecting and Contracting Evaluators

In the Foundation policy records, excerpts detailing selection of evaluators and the particulars of the contracting⁷ process represent 8% of the overall framework. The codes here are divided between two major categories indicated in Table 10. The first contains codes related to administrative details about organizing evaluations (including contracts and proposals and the financial, time, and personnel resources devoted to evaluation). The second category includes information about the selection of particular evaluators including their values, technical expertise, interpersonal skills, and relevant content knowledge. Each of these categories will be discussed in turn.

⁷ An important note to share here is that the Foundation does not use the term “contracting.” In this study, contracting is used as a general term to describe the agreement between the Foundation and the evaluator, which might include information about the resources, time, and personnel dedicated to the project.

Table 10
Codes Pertaining to Selecting and Contracting Evaluators

Selecting and contracting evaluators		
Contracts and Proposals	7	20%
Financial Resources	15	43%
Time	4	11%
Personnel	1	3%
Evaluator selection	8	23%
Evaluator values		
Evaluator technical expertise		
Evaluator interpersonal skills		
Evaluator content knowledge		
Subtotal:	35	

Note. The codes in this domain represent 8% of the overall coding framework.

Contracting evaluations. Seven excerpts from the document analysis were coded with an excerpt indicating discussion of proposals or contracting. These excerpts describe how requests for proposals (RFPs) are developed and the criteria upon which they are evaluated. The excerpts also offer advice to those responding to requests for proposals. According to one excerpt, “In almost all cases, Foundation staff members generally determine the direction of a program evaluation and then open the evaluation for a limited competition.” The program officer is responsible for designing the RFPs (a process discussed in more depth in the following section). The Foundation also offers suggestions to those responding to the request for proposals. Several of these excerpts discuss how evaluators should devote significant attention in the proposal to describing anticipated problems and how the evaluators might address them. This is an operationalization that clearly relates to an assumption discussed in the previous section—specifically, the belief held by the Foundation that evaluation is a complex activity. The policies also suggest that prospective evaluators should be realistic about their timetables, dedicate reasonable space in their proposal to discussing project management issues, and provide evidence that the evaluators will have access to the data they intend to analyze as a part of the evaluation.

Once submitted, these excerpts indicate that proposals are evaluated according to the quality of the proposed research questions, design feasibility, and the potential value when compared to the available budget. According to the policies, the Foundation often consults with external experts to review proposals. One interviewee explained the process further, indicating that when evaluators are contracted, the documentation includes “the scope of work, the proposal, the narrative, the budget, and the budget narrative ... and that gets submitted, and what gets issued is the letter of agreement. The letter of agreement attaches the proposal and becomes the official contract.”

Evaluation resources. Beyond how evaluators are officially attached to projects, the policies also shed light on some of the financial, personnel, and time resources that should be devoted to evaluation efforts. The majority of these codes (75%) describe the financial resources. The Foundation’s overarching policy is that 20% of its grant making dollars are dedicated to research and evaluation. This claim is made consistently in a variety of records. Interestingly, research and evaluation are always mentioned in tandem when this fact is cited; however, these are two very different activities as detailed in other Foundation writings. Research refers to empirical investigations of phenomena related to the ultimate goals of the Foundation, whereas evaluation (the subject of this investigation) is always tied to a program that has received funding from the Foundation. It is never specified in the documents how much of this 20% figure is spent on either activity.

Furthermore, the 20% of grant-making dollars dedicated to research and evaluation is a Foundation-level, rather than initiative-level policy. The funding per evaluation project differs according to the type of evaluation and stage of the program, ranging from “small program assessments to multimillion dollar outcomes evaluations.” Beyond an overall estimate describing how much money is devoted to evaluation, the Foundation also suggests specific items and processes to which monies should be directed. Some of these include allocating money to pay for an

evaluation liaison between the program and the evaluation team, dedicating money specifically to the planning process, incentivizing participation and paying for performance, and meeting specific Foundation requirements. These requirements relate to the expectation the Foundation has that data collected through evaluations should be shared through a large publicly available data-housing web site.

According to the Foundation staff (as indicated through interviews), budgets for evaluation are determined on a team-by-team basis. When asked about how an evaluation budget for a particular program would be set, one senior program officer describes the preferred approach as letting the particular tasks associated with each evaluation determine the budget. However, this officer noted that sometimes, budgets are instead determined in a more top-down fashion by which someone in the leadership will set aside a particular amount for an evaluation, then the evaluation design must take that limit into account. Finally, the eventual goal for a particular evaluation might influence the budget. In one example offered by a senior program officer, an evaluation was designed with the objective of publishing the findings in a government-held clearinghouse with particular threshold criteria. These criteria pre-determined the number of sites that must be examined, which then determined the evaluation budget for that project.

Time. The codes having to do with time devoted to evaluation are diverse. In one case, the excerpt refers to the Foundation's policy to grant no-cost extensions to programs without penalty to ensure that sufficient time is allocated to complete reporting. The other three instances when this code occurs relate to how much time should be devoted to a particular role, allocating sufficient time for data collection and analysis that accounts for unanticipated challenges, and dedicating some time to engage in capacity building activities with stakeholders to educate them about the evaluation and/or the larger issues at hand.

The most important part about contracting evaluations is selecting the evaluator. From the Foundation organizational perspective, once it is determined (by strategic team) that a program should be the object of an evaluation, a Research & Evaluation (R&E) Program Officer is assigned to the project. The R&E program officer is then responsible for commissioning the evaluation, and selecting the evaluator. The processes involved in determining this are further described below.

Evaluator selection. Policies that govern the selection of evaluators to carry out the work represent 23% of the codes in this domain. Of chief importance to the Foundation is the idea that the selected evaluator is independent. This is explained because according to the Foundation, having an independent evaluator enhances the external credibility of an evaluation, and ensures fewer conflicts of interest.

The importance of independence is echoed in the interviews with Foundation staff. The underlying belief is that contracting independent evaluators, and standing behind their results (regardless of whether they are positive or negative) contributes greatly to the Foundation's credibility and reputation for excellence. In Foundation practices, an independent evaluator is defined as someone who is neither employed at the program or with the Foundation. The independence does not, however, extend to the working relationship once the evaluator is contracted. There is a great deal of collaboration between the evaluator, program officers, and program staff as the evaluation takes place. The Foundation staff describe several ways that they promote independence in evaluations. A chief example is transparency about evaluation processes and findings. One program officer provided a description for how the Foundation maintains independence, saying, "There's many ways of establishing independence, ranging from having an independent advisory committee, to having your data be subject to audit, and your methods be subject to review and audit, to having the independence that we generally require, which has more to

do with independence from [the program]. In other words, the program does not directly pay for the evaluation, and we prefer that route to demonstrate independence.”

The choice to have evaluators that are independent by form, rather than function is explained more through interviews. Several program officers discussed the benefits that they see when the evaluators work more collaboratively with program personnel and Foundation staff. These include richer data and better communication that results in greater staff buy-in and support of evaluation findings.

Aside from detailing how the Foundation defines independence, the process involved in contracting evaluators also brings up interesting considerations about whom the Foundation selects to lead its evaluations. There are two primary ways that evaluators are identified to work on Foundation program evaluations. The first is through an invited bid process, by which invitations are sent to a select group of evaluators (typically individuals or firms with which the Foundation has previously worked and had a positive experience) to solicit their interest and thoughts about evaluation design in an RFP. The second is sole-sourcing, a process by which program officers directly contact the person or firm that they would like to conduct the evaluation. There is no clear written guidance about when program officers should choose either method; however, there seems to be an implicit preference for the invited bid process. Based on initial interviews, there seem to be some differences among program officers for their preferred method of contracting evaluators. Despite differences in practice, several program officers related a similar sentiment about the Foundation’s attitude towards sole-sourcing evaluations versus issuing a request for proposal, stating, “[the Foundation] prefers it if we put things out for bid.” In an effort to better facilitate the selection of evaluators, the Foundation has developed an evaluator database with information about

over 100 individual's areas of content knowledge and particular skill sets⁸. However, only two of the program officers interviewed indicated that they had used the database successfully⁹.

Some of the program officers indicate that contracting evaluators through a competitive bid process is always their preference. Explaining this choice, one program officer says, "I have competed [sent out to bid] almost every evaluation, and I've done that because I believe competition actually probably generates the best ideas, and sometimes you get surprised on who kind of surfaces as the strongest team." A contrasting view is offered by another program officer here, "To be perfectly honest, sometimes ... you need it to be done fast, and you know that person has an expertise, so you can talk to a couple of people and see which one seems better, and you can get it happening fast. Putting out a call, you know, the RFP, it takes time. And then you get a strong proposal, but sometimes you choose somebody you don't know at all, and it works out differently that you expect it to."

However, a concern was expressed by one staff member that, "In my view it happens way too often that [we contract] evaluators that were already worked with." This staff member further explained the concern, indicating that sole sourcing and doing invited bids limits the diversity of the evaluation pool. More specifically, the lack of diversity in evaluators is problematic because "in evaluation, you want the evaluator to match the population they are evaluating." This officer pushed the issue further, indicating that the Foundation could go so far as to play a role in promoting the diversity of evaluators in the field generally, saying, "I would love there to be part of the practice saying 'to what extent can we engage new evaluators in our work? Can we encourage diversity in the evaluation pool?'" However, this idea is difficult to implement given the current Foundation practices. This same officer notes, "We would always with the same people over and over and over

⁸ This database is currently being revamped at the Foundation.

⁹ The fact that only two officers have used the database directly is not necessarily an accurate reflection of the extent of its use. When officers ask colleagues for referrals, those referrals might come from a consultation of the database. That is to say, indirect use of the database is likely more widespread.

again. 'That's something that concerns me.' The desire to work with someone who already has a proven track record performing with the Foundation conflicts with this expressed interest in increasing the diversity of the evaluator pool.

Despite this concern, there are significant policy-related reasons that direct program officers to sole-sourcing evaluators. As stated above, there is a policy that indicates that the starts of evaluations should coincide with the start of programs. Putting evaluations out to competitive bid takes time, and results in a delayed evaluation start. This might lead to a situation in which evaluation results are delayed so much that they are not useful. Also, there is less risk in choosing a known quantity (an evaluator whose work quality is already familiar to you). Program officers can feel more assured of the quality of products they can expect if they are working with someone they already know. This is particularly important given the fact that the Foundation is interested in gathering credible evidence to inform policies on a national stage. As one program officer succinctly put, "this isn't amateur hour."

These ideas provide an example where tensions arise between particular operationalized policies. The concern is raised about the time it takes to go through the RFP process because another of the Foundation's implied policies is that the evaluation should begin when the program begins (noted above). This problem is exemplified when a program officer says, "And part of the challenge is, if I get a task to do an evaluation, I have a choice of picking a single evaluator, just by asking people around here, 'do you know anybody who could do X, Y, Z?' and then going with that person, versus saying, 'I am going to issue an RFP and it is going to take me 3-6 months to identify someone to get started on an evaluation.' Which would you choose?" These tensions are, at present, unresolved within the Foundation. There are beliefs held among individual program officers that suggest mild discomfort with how things are presently done, but there is no active effort to change it, or mechanism by which it might be changed. This idea will be revisited later in this study.

Regardless of how the evaluator is contracted, there are specific qualifications sought in the people who undertake this work. Some of the desired evaluator qualities indicated among these excerpts include that an evaluator should have a track record for balancing being on-time in completion of deliverables, yet being flexible and adaptive to contextual situations. Beyond that, there is little discussion of evaluator characteristics. In fact, there are several available codes in the evaluation policy coding framework that do not appear in the Foundation records. There are no specific mentions about the evaluator's particular skills relating to their technical expertise, content knowledge, interpersonal skills, or values (all of which are discussed with frequency in the literature about evaluator competencies (Stevhan, L., King, J., Ghore, G., & Minnema, J., 2005) In interviews, however, some of these qualities were discussed in very broad terms. When contracting evaluators, program officers look for relevant content area expertise, "rigor of methodology," feasibility of proposed work, "clarity of communication," and general working style. One R&E program officer, in speaking about a particular evaluation, said, "I think we definitely are trying to pick people that know more about this (the object of the evaluation) than we do for sure ... You usually pick someone because you trust their judgment and you think they're really going to have better ideas than you do."

The lack of more deliberate discussion in the Foundation records of evaluator background characteristics should not, however, lead to a conclusion that the Foundation has no specific objectives or requirements in mind when contracting evaluators. There are several other reasons that might explain why these areas are missing from the records. One hypothesis is that the requirements for an evaluator are so obvious as to not necessitate explicit articulation. Another explanation might be that the desired evaluator qualities might vary so much from program to program that over codifying it in the records that describe the evaluation policies might hamstring the selection process and prohibit identifying an evaluator who is the best fit for a specific task. Finally, policies governing

these aspects of evaluation context might be communicated in more indirect ways (for example in requests for proposals (RFPs) or in the language of particular contracts) and thus make up more of the informal policies that govern evaluation.

The next step in the evaluation process is to design and conduct the evaluation. The processes by which the Foundation undertakes and supervises these efforts are discussed next.

Designing and Conducting the Evaluation

Codes pertaining to designing and conducting evaluations are the most well-represented in the overall framework—20% of the codes lie in this domain. The specific categories nested within this idea provide a chronological walk through the process of conducting an evaluation. Steps necessary to understand the program begin the process, followed by designing the evaluation. Next are methods chosen in the evaluation and the data collection and analysis procedures. Once those are complete, findings are derived, and conclusions are developed based on those findings. Finally, ethics when conducting evaluations are discussed.

Table 11
Codes Pertaining to Designing and Conducting Evaluations

Designing and conducting the evaluation		
Understanding the program	14	16%
Designing the evaluation	43	49%
Methods	8	9%
Data collection	9	10%
Data analysis	7	8%
Findings	3	3%
Developing conclusions	2	2%
Ethics	1	1%
Subtotal:	87	

Note. The codes in this domain represent 20% of the overall coding framework.

Understanding the program. There are 14 instances in which this code is applied to evaluation policy excerpts. The Foundation maintains that understanding the program is a critical initial step in the evaluation process. Some of the excerpts indicated here detail why the Foundation

believes this, while others describe important parts of the process of understanding the program.

Taking steps to understand the program is important because it helps to ensure that the evaluation is asking questions to which the answers are not yet known, and that the outcomes being measured in the evaluation are reasonably expected to result from program activities.

The Foundation policy records recommend several processes for understanding the program, including suggesting what, exactly, should be found out about the program (how and why the program started, the assumptions that underlie its activities, and how those activities connect to intended outcomes), and the information sources that could be consulted to figure out that information (grant documents, board reports, research articles, etc.). The policies indicate that another important part of this process is to uncover some of the assumptions about the program held by Foundation and program staff. There is also allowance for the evaluator to introduce his/her own opinions about what the program is supposed to do. Finally, the policies indicate that the information gathered in this process should be included in the final evaluation report.

Many of the excerpts in this category (50% of them) specifically recommend undergoing the process of developing a logic model as an important aspect of understanding the program. A logic model is described in these excerpts as a visual depiction of program inputs, activities, outcomes (short-, medium-, and long-term) and the logical connections between those elements. One of the internal videos is entirely devoted to educating the Foundation staff about logic model development and uses. The Foundation's policies also indicate that the process of developing a logic model should include stakeholders both to increase clarity about the program and reach a common understanding, and because it can lead to important conversations about what should be evaluated and the development of evaluation questions. The idea of developing logic models was also mentioned several times during the interviews (with both research and evaluation and program team members) as an important part of understanding the program.

Interviews with Foundation staff, for the most part, did not shed additional light on this area. One exception was one program officer indicating that based on his/her previous experiences, s/he now always includes a 3-6 month planning process in evaluation contracts, partly to ensure adequate time for evaluators to familiarize themselves with the program and its surrounding context. The lack of discussion around this particular topic may lead to a conclusion that R&E program officers do not play an active role in this part of the evaluation process—rather, this task falls more squarely in the domain of the evaluator.

Designing the evaluation. Excerpts describing the steps involved, and some of the key components associated with designing an evaluation make up 49% of the codes in this domain. Within these excerpts, there is great emphasis placed on developing evaluation questions (40% of the codes within this subcategory relate to this topic). This emphasis seems partly due to an underlying assumption held by the Foundation that there is a strong relationship between developing good evaluation questions and the eventual use of evaluation findings. This is exemplified in an excerpt that states, “good questions, when they are thoughtful and well-informed given the range of perspectives that went into developing them, are more likely to yield findings that are useful, relevant, and credible.” Furthermore, the Foundation explains “high-quality, well-informed questions ground the evaluation in a way that facilitates subsequent phases of the evaluation.”

Beyond explaining how good questions promote use of evaluation findings, the policies in this area describe the qualities of good evaluation questions, and the steps that evaluators should take in developing the questions. The Foundation policies state that good evaluation questions reflect myriad stakeholder views and values, establish boundaries and scope of work, and ground the process in a way that lays the foundation for subsequent evaluation phases. The excerpts go on to suggest ways of accomplishing this: evaluators should involve a range of stakeholders in the question

development process, call on external sources for information, provide opportunities for people to offer input, identify themes that surface in meetings unrelated to evaluation, but attended by stakeholders, and finally, prioritize questions according to “who needs the information, when, and for what purposes.”

The interviews, however, share a different depiction of how the question development process unfolds. Generally, research and evaluation program officers initiate the evaluation design process. R&E program officers, occasionally in consultation with program officers, think through developing research questions before they work on identifying an evaluator to do the work. When an RFP is developed, it usually contains at least a draft of the research questions and suggestions for answering them. As described in the interviews, these topics are then refined through a collaborative, “back-and-forth” process. This idea is illustrated in this quote, “Having an R&E officer involved in the program design is critical. And it helps the program designers and the team decide, well, what are our objectives for this program? And if those are our objectives, how are we going to institute data collection to answer that question.” The necessity of constant revision of evaluation questions is captured nicely as a program officer states, “We [R&E program officers] develop the initial questions because people need a guidepost, but I have found in every single case of evaluation that those evaluation proposals totally need to change because the programs look completely and totally different than what we thought they would.”

Beyond developing evaluation questions, other aspects of designing the evaluation indicated among these excerpts include anticipating evaluation challenges, exploring available data sets, aligning evaluation to the program’s stage of development, determining logistics (including scheduling and workflow), and developing a plan for reporting and disseminating findings.

Methods, data collection, and data analysis. Determining appropriate methods to answer research questions happens in concert with designing the evaluation. Taken together, methods, data

collection, and data analysis represent the bulk of work that evaluators do. Excerpts tagged with these codes represent 27% of the codes in this domain (when broken down, excerpts detailing policies that describe methods make up 9% of the codes, data collection excerpts are 10% of the codes, and data analysis excerpts are 8%). The policy records discuss these concepts in general, rather than specific ways. These are not necessarily prescriptive policies, rather, they provide general guidelines and speak to a particular philosophy about using data to answer evaluation questions.

When it comes to choice and use of methods, the Foundation's policies suggest that the methods chosen to evaluate programs should involve "disciplined inquiry" that uses "approved scientific procedures." The policies also indicate that the methods chosen should be rigorous and made transparent to eventual readers of an evaluation report. In three of the excerpts, a particular approach is mentioned—the randomized experiment (commonly known as a randomized control trial or RCT). This approach is referred to as the most credible type of outcome study. However, the Foundation policies also indicate that there are contextual factors associated with evaluations that may prohibit the use of randomization. In those instances, the excerpts indicate that other approaches may be used, but the approach should be geared towards reducing bias in findings. This suggests that a randomized experiment is the preferred approach to evaluation when conditions support it.

These ideas were echoed in the interviews. Several times, RCTs were touted as the preferred methodological approach. However, several people acknowledged that sometimes the program context interferes with the successful execution of an RCT, and may lend better to a different design. One program officer summed up the situation nicely, saying "I think we're trying to design it in response of the circumstances that are before us in a particular strategy." It also seems to be the case that people have particular methodological preferences based on their training prior to coming to the Foundation, exemplified in the following statement, "I like rigorous scientific research

designs. So, I don't really like these process, soft, kind of subjective evaluations ... I'm not trained that way." Another factor that sometimes influences design choices and methods is the intended audience of evaluation findings. The Foundation may target a particular number of sites or a certain approach in order to influence a particular group, like policy makers who may attach more credibility to certain evaluation designs.

Perhaps unsurprisingly, data collection and analysis codes are frequently applied to the same excerpt, as these aspects of evaluation are most often discussed in concert with one another. In many cases, the guidance offered by the policies in this area describe approaches to using extant data sets to answer evaluation questions. Here, the Foundation indicates that evaluators should explain their access to and familiarity with these data sets along with piloting of some of the analysis. Another suggestion made in these policies is that evaluators carefully document the steps that they take in collecting and analyzing data. The Foundation also indicates that there are situations in which requesting money to pilot analysis techniques is appropriate. Only one of the excerpts tagged with these codes describes a situation in which the evaluator will be collecting primary data. All that is said from a policy standpoint is that the collection methods and analysis plan are appropriate and that they match the design.

The interviewees also shared sparse information on this topic. It seems that once the evaluation is designed, and data collection and analysis plans are determined, the R&E officers have more of a "hands-off" approach. During this phase of the evaluation, R&E officers check in less with the evaluators on their project because they are just "doing their thing." The distance between R&E officer and evaluator during this time also plays into the Foundation's desire for independent evaluations. The Foundation does not interfere with these activities. Data collection and analysis are firmly in the domain of the evaluator.

Findings and developing conclusions. Once data is collected and analyzed, findings are reported and conclusions are developed. There are only five codes (5% of codes in this domain) that describe these processes. Most of these excerpts serve to characterize what the Foundation considers ideal findings—claims that are well supported by evidence, conclusions that fit analysis, and descriptions that include enough information such that readers can draw inferences. Furthermore, findings should be constructive, impartial, useful, relevant, and credible. The policies also indicate that it may be appropriate for evaluators to make recommendations at the conclusion of a project. According to the policies, these recommendations should demonstrate insight, flow from the findings, and be practical and realistically executable.

Ethics. Ethics is a topic that is not explicitly discussed much in either the Foundation records or through the interviews. There are two instances in which the ethics code is applied. One of these cases indicates a policy that the evaluation design is an ethical one that includes human subject protection safeguards. The other instance is an indication that involving stakeholders in the evaluation constitutes an ethical practice.

The next two sections, administration and stakeholder involvement, disrupt the chronology of the discussion that has taken place thus far. These processes are present through all the aspects of evaluation practice. The excerpts indicated here discuss logistical aspects of conducting evaluations, and the ways in which relevant parties are involved in the processes.

Administration

The codes that fall under the administration domain mostly pertain to information about particular responsibilities and relationships among different evaluation participants. These include general information about evaluation roles, followed by specific roles for the Foundation, program, and evaluators. These are supported by policies relating to communication, managing relationships, and administration. These concepts are represented in Table 12.

Table 12
Codes Pertaining to Administration of Evaluations

Administration		
Roles	9	15%
Foundation's role	5	8%
Program's role	7	11%
Evaluator's Role	10	16%
Communication	26	42%
Managing relationships	14	23%
Administration	8	13%
Subtotal:	62	

Note. The codes in this domain represent 19% of the overall coding framework.

Roles. The concept of roles appears several times in this coding framework. Generally, these excerpts refer to statements that describe two different areas. The first is details the roles that people will take *during* the evaluation. The second is describing the roles people take in *planning and preparing* for the evaluation. For example, an excerpt that describes the first category, states that a proposal should include, “detailed information regarding how the project will be organized ... which employee is responsible for assuring adherence to project schedules, monitoring expenditures, and addressing delays.” Other themes that emerge in discussions of roles to be assumed during an evaluation include establishing decision-making responsibilities and supervision structures that take prior relationships among parties into account. The second type of excerpt explained above includes quotes like, “While [the Foundation] usually develops initial evaluation plans in consultation with the director of a funded program and key stakeholders, the selected evaluators heavily influence the final evaluation design.” This quote is indicative of all the excerpts that fall into this category—they all have to do with the roles people assume in designing the evaluation and developing questions.

Various responsibilities. The Foundation policies also specify the specific roles and responsibilities ascribed to the major players in an evaluation: the evaluator, the program, and the Foundation. If these three categories are collapsed in the coding, ascribing particular responsibilities to different evaluation parties accounts for 35% of the codes in this domain.

As expected, most of the codes across these three subcategories relate to the evaluator's role in conducting evaluations. According to the policies excerpted here, the evaluators work with Foundation staff to determine the evaluation questions and research design, but are ultimately responsible for setting the mission and vision of the project. Evaluators are expected to be able to "see both the forest and the trees." Furthermore, in these excerpts, evaluators are also encouraged to take a hands-on approach to conducting evaluations as suggested by phrases like, "Do the hard work—dig in up to your elbows if that's what it takes" and "do the heavy lifting." In a couple of these excerpts, the point is made that once evaluation work is underway, the Foundation refrains from interfering with the evaluation. As previously mentioned, the Foundation places great importance on the idea that evaluations are to be conducted independently.

Those responsible for the overseeing the program being evaluated are also expected to take an active role in assisting the evaluation. First, program directors are seen as stakeholders who can make valuable contributions to designing the evaluation. Those responsible for program operations are also recognized as key gatekeepers that can facilitate access to necessary data. The Foundation encourages program personnel to play an active part in the roll-out of the evaluation to set the expectation to those working at the program that this is an important process as a means to make the evaluation activities happen smoothly. Another point of emphasis in these policies is that the program will, by necessity, play an important role in data collection activities. Program personnel are expected to provide the evaluators with necessary data to support the evaluation, and in some cases, to assist with primary data collection activities.

Finally, the role that the Foundation officers play in an evaluation is described in five of these excerpts. The Foundation's policy is that program officers are involved initially in early evaluation design activities (and in concert with the evaluator once s/he is contracted). Then, these policy statements reiterate the point that the Foundation program officers are not to interfere with

any of the evaluation activities once the bulk of the evaluation work is underway, and particularly as conclusions are being reached and findings are being shared.

When the idea of roles came up during interviews, a theme that emerged consistently is that the process of undertaking evaluations is very collaborative, particularly at the beginning of the process (when questions are being developed) and at the end (as findings are being shared). The policies that describe these more structural indications of the roles and responsibilities of various parties involved in evaluation activities are supported by policies that detail how communication should take place between the parties and how the various relationships should be managed.

Communication. Thirty-three percent of the codes in the category of conducting the evaluation describe the communication that takes place between parties as the evaluation is ongoing. These excerpts fall generally in to three categories: those that describe the purpose of ongoing communication, specific strategies to support open communication, and the organizational processes that underlie communication efforts. Generally, the excerpts that describe the purpose and provide justification for some of the specific communication strategies indicate that greater frequency and higher quality communication between program personnel and evaluators will help create a smoother and more effective evaluation process.

The policies here also outline several strategies that evaluators may adopt to support the Foundation's idea of effective communication. These range in scale from ideas like conducting an ice-breaker activity at the first meeting, to consistently adopting facilitation processes that allow all voices to be heard. Other themes to emerge in this section of the coding include managing stakeholder expectations for evaluation, bringing clarity to murky situations, and developing a work plan to support communication. Finally, indications of organizational process that emerge through this analysis include a description of the channels of communication, the venues through which

communication might take place (phone calls, email, in-person meetings), and an identification of particular occurrences that should trigger a communication between parties.

According to interview participants, communication plans are developed on a case-by-case basis for each evaluation. Program officers consistently reported that there is more communication at the beginning and ends of projects as that is when more decisions are being made. Once evaluation activities are well underway, the communication lessens somewhat. Most program officers indicate that they schedule a monthly call with the team at this point, but that the call might be cancelled when there is nothing to talk about. Beyond phone calls, occasional in-person meetings were reported in addition to requiring written quarterly reports summarizing activities.

Managing relationships. A concept related to communication found in these excerpts is the idea of managing relationships (which appears in 23% of the codes here). In fact, 57% of the excerpts tagged with this code are also tagged with communication. The Foundation outlines policies that identify the particular parties that should have relationships with one another, the desired characteristics of those relationships, and tools that might facilitate building and managing those relationships. In addition to the more obvious relationships that are needed to support a successful evaluation (between evaluator and program operator, for example), the policies here also suggest that the Foundation has an interest in helping to promote relationships between stakeholders such that stronger networks can be built in support of a particular intervention or issue. The Foundation wants to promote relationships that are characterized by trust, mutual respect, and are reciprocal. The policies further indicate that relationships like this are developed through having an equitable balance of power, determining mechanisms for sharing timely feedback about the program, formalizing work plans and communication protocols, and establishing clear procedures for resolving misunderstandings that might arise.

Administration. Beyond the specific tasks devoted to promoting communication and developing and maintaining productive relationships, there are a few policies that speak to more general administrative tasks that occur during an evaluation. The areas that these excerpts address are diverse and include policies to guide evaluators such as setting a mission and vision for the project, establishing flexible work plans, managing work flow, monitoring subcontractors, and ensuring timely and appropriate completion of deliverables.

Stakeholders

Like administration, stakeholder involvement represents another domain that overarches several of the processes involved in conducting evaluations. As discussed earlier, the Foundation emphasizes involving stakeholders in the evaluation process because of the assumption that their buy-in to the process will promote the eventual use of evaluation findings. The ways in which the evaluation policies describe stakeholders can be divided into several categories: identifying stakeholders to be involved, offering reasons for their inclusion, describing specific ways they can participate, and building their capacity for participating in evaluation. These are represented in Table 13.

Table 13
Codes Pertaining to Stakeholder Involvement in Evaluations

Stakeholders			
	Identifying Stakeholders	12	15%
	Stakeholder Inclusion Reason	26	33%
	Stakeholder Participation	33	42%
	Capacity Building	7	9%
Subtotal:		78	

Note. The codes in this domain represent 18% of the overall coding framework.

Identifying stakeholders. The codes that pertain to stakeholders are well-represented in this framework. This domain accounts for 18% of the overall codes. The majority of these codes come from one document detailing stakeholders involvement in evaluations. Stakeholders are broadly defined as “those who have a vested interest in that which is being evaluated, and thus,

would be in a position to use the evaluation results in some way” and as “those who have some knowledge of, role in, or relationship to the program being evaluated.” Additionally, according to the Foundation, stakeholders are identified both in terms of their roles and responsibilities as it relates to the program, and according to their backgrounds, attitudes, and beliefs.

As it comes to roles and responsibilities, generally the Foundation advocates including stakeholders that have some decision-making power. These might be program managers, program developers, or those who are in some way responsible for the initiative’s success. The policies also indicate that the pool of involved stakeholders should be representative of various levels of program function beyond those directly responsible (including recipients of services, service providers, and policymakers). Aside from the particular function a stakeholder serves in relationship to the program, the Foundation’s policies also indicate that the process of stakeholder selection should be attentive to the attitudes, beliefs and knowledge of individuals. Some specifics include individuals who provide buy-in and support for evaluation activities and those with great interest in the issues being examined. Finally, the policies value incorporating diverse perspectives and cultural, religious, ethnic, and geographical backgrounds.

Stakeholder inclusion reason. Another concept that makes up this category is the justification offered for including various stakeholders. This concept was noted 26 times in the documents. There are frequent incidences of co-occurrence of the concept of stakeholder identification and the reasons particular stakeholders are to be included. Six of the excerpts coded with the identifying stakeholders code were also coded with inclusion reason (50%). Many of these inclusion reasons are tied to use. The Foundation advocates including people in influential roles because they are more likely in a position to be able to act upon recommendations made in evaluations. On the other hand, the Foundation suggests that bringing in people with diverse perspectives and from a variety of roles within and outside of the organization will result in

recommendations that “are more likely to be accepted by a broader constituency and implemented more fully and with less resistance.”

Other reasons offered by the Foundation for including stakeholders in evaluation include promoting the use of evaluation findings (27% of codes within this subcategory), improving the process of conducting the evaluation (23%), developing better evaluation questions (19%), building networks (12%), gathering support for evaluation activities (8%), developing findings that are credible (8%), and building capacity of stakeholders to conduct their own evaluations (4%).

Stakeholder participation. Excerpts tagged with this code specifically describe strategies for involving stakeholders in the evaluation process, and identify specific activities in which stakeholders should be engaged. A distinguishing aspect of the excerpts in this category is that they span five different documents. Furthermore, there are a variety of different audiences represented—evaluators, Foundation personnel, and grantees. It is clear that including stakeholders in evaluation is a very important part of the Foundation’s evaluation policies.

There is greater emphasis placed on involving stakeholders in the early stages of the process, particularly when it comes to understanding the program being evaluated, developing appropriate questions to guide the evaluation activities, and designing the evaluation. Together, these three themes appear in 89% of the excerpts that were tagged with more than one code. These excerpts also indicate that stakeholders should be involved in data collection. This idea is represented in the other 11% of the excerpts.

Involving stakeholders in the process of understanding the program is advocated by the Foundation because of an underlying assumption that those with more direct experience are better informants about program operations. An added benefit of involving stakeholders in this process is that collaboration between evaluators and stakeholders in describing the program might in fact serve to strengthen and solidify a common program understanding among all relevant groups. Once an

understanding of the program is reached, the next step in the evaluation process is to develop questions and design the evaluation. The Foundation policies indicate that involving stakeholders in these processes results in better evaluations—defined as reflecting a wide range of perspectives and yielding appropriate information that will ultimately be useful in informing program practices and progress toward Foundation objectives. In the two instances in which involving stakeholders in data collection activities is described, the key feature seems to be gaining access to necessary data, which can often present a challenge without the cooperation of a key gatekeeper.

The instances in which the stakeholder participation code was exclusively applied describe key engagement strategies. These range from more general ideas (open communication between the evaluator and stakeholders throughout the evaluation process) to more specific approaches (providing something concrete to which stakeholders must respond as a way to ensure dialogue continues). Furthermore, some of the excerpts in this subcategory describe the necessity of ensuring the identification of appropriate stakeholders for particular tasks, and ensuring their representativeness of the program and its intended recipients.

The responses to interview questions regarding stakeholders do not strongly align with what appears in the policy records. Engaging stakeholders in the evaluation process was mentioned three times by participants, and in all cases, it specifically relates to stakeholder involvement at the end of the evaluation, when findings are being prepared and disseminated. The participants indicate that engaging program personnel is critical at this stage for several reasons. First, there is an underlying belief that it is ethical to share findings with the program before any other audience. Second, some of the program officers believe that stakeholders can play a very important role in finalizing what the evaluation products look like. Finally, program personnel can assist the Foundation with developing evaluation products that are most appropriate for different audiences.

Reporting Findings

Once evaluation activities are complete, the Foundation turns its attention to sharing findings, represented in Table 14. This is an extremely important process for the Foundation because it directly ties into its goals of informing policy makers and practitioners. The coding framework for this domain includes several categories: the characteristics of findings, the intended report audience, reporting format, and the content of reports.

Table 14
Codes Pertaining to Reporting Evaluation Findings

Reporting findings		
	Findings Characteristics	8
	Audience	13
	Format	11
	Content	11
	Other	2
Subtotal:		24

Note. The codes in this domain represent 11% of the overall coding framework.

The Foundation engages in several types of reporting across a variety of levels. The most directly related to the topic of study here is evaluation reporting that happens on the program level. The Foundation also engages in annual reporting, during which it compiles information gathered through evaluation efforts across programs. The various types of reporting that happen demand different attention to audience, format, and contents. The policies described here indicate to whom and how particular aspects of the evaluation should be shared. There are several components of interest here: specifically, the qualities of the findings; the intended audience with whom the evaluation specifies sharing findings; the report formatting specifications; and the content that should be shared. More so than in the other categories in this domain, the policies governing reporting are explicit, rather than general.

Findings characteristics. The eight excerpts that describe characteristics of the evaluation findings describe both qualities and the substance of the information. According to the Foundation,

of primary importance is “providing high-quality, objective information.” Furthermore, evaluations are designed to provide information that indicates the effectiveness of the program. Specifically, the documents detail, “capturing the factors that account for our successes and failures,” “measuring the extent to which, and the ways in which, a program or initiative’s goals are being met, and how the program or initiative might be contributing to the organization’s mission,” and “understand[ing] the perceived impact of its work on all affected parties.” These three statements create a narrative of what evaluation results should look like. First, results are supposed to document the successes and challenges of funded programs. Second results measure the magnitude of these effects in some way. Finally, results assess what those effects mean for the organization and other related parties.

Audience. The Foundation identifies audiences for evaluation findings that are both internal and external to Foundation personnel and staff of Foundation-funded projects. Internally, the policies identify program officers, management, the Board of Trustees, and key program stakeholders. Externally, the Foundation casts a wide net while identifying potential audience members for evaluation findings, including public and private decision-makers, academics, researchers and scholars working in similar areas, the general public, and stakeholders in other programs. This was echoed in the interviews with key Foundation staff, and the selection of audiences is explained further. It is closely tied to the goal of use of evaluation findings. The Foundation wants the findings from its programs to spur certain actions. This, then, informs with whom these findings are shared.

Format. The desired format of evaluation reports is also detailed in the evaluation policies, and differs according to whether or not the reporting is at the program or Foundation level. Some of the excerpts to which this code is applied relate to some of the qualities of evaluation reports, and others discuss the venues in which they are reported. A primary indication in these excerpts is that evaluation reports should be formatted in an understandable way as the intention is to widely share

evaluation findings. To enhance understandability, a basic report format is detailed, which includes an executive summary, a brief and clear body, and appendix that includes the more technical details.

The desire of the Foundation to share evaluation findings in academic, peer-reviewed journals also dictates some formatting issues—mainly that the report conforms to the conventions of the academic journal of interest. The Foundation also shares evaluation findings through its website. Other formats mentioned in these excerpts are interim evaluation reports, used to share useful information with program stakeholders, and the annual scorecard, which shares Foundation lessons across content areas.

Content. As with reporting formatting, the desired content of reports differs between program-specific evaluation reporting and reporting on larger Foundation objectives. In the case of the latter, the policies indicate that reporting should include information that speaks to the health of program development, the services to grantees, and the Foundation’s impact on its areas of interest.

Regarding program-level evaluation reporting, the Foundation policies include a very specific and detailed outline of an ideal report. This includes information about the object of the evaluation, the reason for the evaluation, the intended audience of the evaluation, the evaluation questions, the key findings, the conclusions, and the recommendations. Other areas that the policies emphasize as critical components of an evaluation reports are the technical details (including methods and challenges) and the evaluation limitations. Finally, some of the desired qualities of evaluation reports are discussed including accuracy, balance, demonstration of content knowledge, clarity, brevity, responsiveness, and treatment of biases.

Other. Finally, two excerpts about reporting indicate evaluation policies that do not fall into any of these categories. Both of these have to do with a policy that when primary data is collected under the auspices of a Foundation-funded evaluation, that data must be shared with an online social research consortium so that it may be used by other researchers interested in similar issues.

During the interviews, another type of internal reporting was discussed. Specifically, at the conclusion of each program and evaluation, a Program Results Report is developed. The process for developing these involves contracted writers working with relevant program documents and conducting interviews with involved parties. The results reports provide an overall description of what happened in the program and what was found out about it. These are shared on the Foundation website (and as such, are publicly available). Also, when situations permit, results reports create grant making lessons that are for internal use only. There are two people not directly involved in evaluations at the Foundation involved in this task. They develop these lessons based on interviews with all program participants and a careful review of evaluation products. The lessons are distributed to those directly involved in the evaluation (not widespread in Foundation), but are available to anyone internally.

Evaluation Policy Communication

To address the question of how Foundation evaluation policies are communicated to people both within and outside the foundation, both the audience and format of each of the evaluation policy records presented in Table 1 (which originally appears on p. 20, and is repeated here for convenience) are considered.

Table I

Record Descriptive Information

Record	Excerpts	Codes	Audience	Type	Topic
A	112	158	Evaluators	Document	Stakeholders
B	63	109	Public	Document	Operations
C	34	46	Public	Document	Policy
D	17	22	Public	Document	Policy
E	12	15	Grantees	Document	Policy
F	8	9	Internal	Video	Policy
G	8	10	Internal	Video	Policy
H	7	7	Internal	Video	Policy
I	7	8	Evaluators	Document	Reporting
J	6	6	Internal	Video	Operations
K	4	3	Internal	Video	Operations
L	4	7	Internal	Video	Policy
M	4	5	Internal	Video	Policy
N	3	5	Internal	Document	Operations
O	3	2	Internal	Video	Operations
P	3	3	Internal	Video	Operations
Q	3	4	Internal	Video	Reporting
R	2	1	Internal	Video	Policy
S	1	4	Internal	Document	Operations
T	1	1	Internal	Video	Operations
Total:	302	425			

All of the videos are directed internally (60% of entire sample)—they are instructional videos developed by the research and evaluation leadership for the express purpose of informing Foundation employees about evaluation practices and procedures. The documents represented here (which comprise 40% of the remaining sample), are directed at a variety of audiences. Three of the documents are designed for the general public (37.5%). Two are internal to the Foundation (25%), two are directed at evaluators (25%), and one is specifically for grantees (12.5%)¹⁰.

When examining the coding applications according to the document's intended audience, 42% of the codes (38% of the excerpts) are included in documents written with a public audience in mind. Comparably, 39% of both codes and excerpts are specifically for evaluators (these all come

¹⁰ These determinations were made based on several factors: the title of the document, the availability of the document, the content of the document (particularly in terms of to whom the writing is addressed).

from one document that pertains to stakeholder involvement). Nineteen percent of the codes (sixteen percent of the excerpts) are directed internally. Finally, 4% of both codes and excerpts are included in documents that are written with grantees mainly in mind. An important note here is that these categories are not mutually exclusive. Many of these records are available publicly, so the intended audience might differ considerably from the actual audience. A match is only assured in the case of internal records—these are not made publicly available, so the only possible audience is those who work at the Foundation.

Another way to consider the audience of these records is in light of their primary topic. While all of these records include explicit evaluation policy information, some have different primary foci, which include operations, reporting, and stakeholders. These broad distinctions were made with a holistic view of each record's main goal or premise. Operational records are those that relate directly to Foundation practices (for example, one of the records in this category is entitled, “What is a program logic model” and includes information detailing the process and rationale for developing program logic models). Those records that relate to reporting include information that shares how research and evaluation findings should be communicated and disseminated (as an example, one record is entitled, “Guidance on evaluation reports...” and explains reporting requirements and criteria for funded research and evaluation projects). The category of stakeholders was applied to one document (5%) that was specifically designed to offer advice about engaging stakeholders in the evaluation process. The remaining documents all focus most directly on evaluation policy.

An examination of the primary topic by record audience reveals that there are no instances in which all the topics (operations, stakeholders, reporting, and general policy) are communicated to all audiences. Internally, three of the areas are represented (everything but stakeholders). Policies relating to operations (in addition to general policy statements) are communicated to the public, and contracted evaluators are exposed to policy information that governs reporting and stakeholder

inclusion. Finally, only records that speak generally about evaluation policies are shared with grantees.

Many of these distinctions make logical sense. For example, the information directed towards grantees is broad, and about policies in general. Grantees, who are subject to the procedures involved in evaluation, do not necessarily need to know about specific reporting requirements or operationalizations of Foundation practices. The fact that information about stakeholder participation is not directly communicated to grantees is interesting as they are, in most cases, considered the primary stakeholders of any evaluation. What seems to be happening is that information about the stakeholders' involvement in evaluation activities is brokered through either evaluators or the Foundation staff—it is not intentionally communicated via written methods to the stakeholders directly.

The information about policies that is specifically communicated to evaluators primarily discusses stakeholders and reporting requirements. Of note is that evaluators are the only group to whom information about stakeholders is communicated. This further supports the idea that involving stakeholders in the process is primarily a task that resides with the evaluator. Also, reporting requirements are directly shared with the evaluators because various ways of sharing information are a direct deliverable and expectation of agreeing to work for the Foundation.

Information about policies and operations are shared with the general public. Again, this is logical because the specific tasks associated with involving stakeholders and reporting information are likely not of interest to a broad audience. Instead, through these records, the Foundation is sharing more general information about its practices in an effort to be transparent.

Finally, the Foundation communicates some of its evaluation policies internally. All the topics are represented here, with the exception of involving stakeholders in the process. Obviously, those working at the Foundation have access to (and are likely aware of) the document that

describes stakeholder involvement. However, it is interesting to note that despite the importance of this topic, there is no information about it directly communicated to Foundation employees. This might account for some of the inconsistent alignment found in the analysis comparing record analysis about stakeholder involvement and the interviews with Foundation staff.

It bears reminding that generally, the Foundation does not make an attempt to comprehensively articulate its evaluation policies. This happens in a more casual way overall. If policies were to be more specifically articulated, the Foundation might want to also devote some attention to the audiences with whom particular information should be shared.

The Development of Evaluation Policies

The final question in this section asks about the process by which the evaluation policies are developed. This proved extremely difficult to answer given that many Foundation employees generally believe that the Foundation does not have explicit evaluation policies. The policies seem to manifest in mostly informal ways, and the process they undergo to become common practice is unclear. Some instances in which policies seemed to have shifted over time, however, are noted. This sheds some light on how policy development might unfold. In total, changes in policies were mentioned 16 times in the interviews.

There are two particular shifts mentioned several times that show an evolution in the Foundation's policies. The first is in the degree to which evaluators collaborate with program and Foundation personnel during the process of an evaluation. The second relates to the timing of evaluations as they relate to the start of program activities. Stories told during interviews that shed light on each of these are shared because this is the best place to direct attention in seeking examples of the process behind the development of policies.

At present, evaluations are a collaborative effort between evaluators, Foundation Program Officers, and Foundation R&E Officers, but this was not always the case. Although no one could

pinpoint the exact moment the shift occurred, the general sense was that evaluations have moved to being more collaborative in the past several years. Previously, there was an insistence on a “firewall” between the evaluators and the program from a department director. The thinking was that the evaluation should happen entirely separately from the program activities and personnel. This became problematic because, according to one program officer, “we are so interdependent to understand what the program is, to understand what is a reasonable expectation ... and certainly the ethical idea in the [American Evaluation Association] Guiding Principles for Evaluations states that all the stakeholders must be consulted.” These challenges experienced by those working at the Foundation led to the current model of collaboration.

Further evidence of how the thinking at the Foundation evolved is demonstrated in this quote from a program officer, “It’s primitive ... to think that if you talk to an evaluator at all, you’re contaminating the experiment. So, I think you can learn from the evaluator and they can learn from you and you also need to make sure they’re doing a good job.” However, one of the program officers noted that maintaining a balance between independence and collaboration can be a challenge, noting, “If I want to be really good at objective neutral evaluation, I should be standing back and doing my thing, almost divorced completely from the process. I think we kind of struggle with that.”

The other example of a shift in policy over time relates to the scheduling of evaluations¹¹. Initially, the Foundation used to fund evaluations to see if a program ‘worked’ as program activities were drawing to a close. One program officer explains, “the earlier model was done more at the end. It had a certain elegance to it because you get a lot of benefit from hindsight.” This ties into the

¹¹ This shift in evaluation approach is also reflective of leadership changes at the Foundation, coupled with changing economic circumstances. Howard Freedman, in his role as an early advisor to the Foundation, recommended that evaluation activities begin in concert with program activities. This became less of a priority when he left the foundation and new leadership members assumed responsibility for steering the direction of evaluations. Furthermore, in the economic boom of the 90s and 00s, the Foundation needed to get “money out the door” to programs quickly in a way that exceeded their capacity for evaluation at that time. Because of this, evaluations were frequently commissioned as programs were drawing to their close.

thoughts underlying the other early example about evaluator independence. Early in the Foundation's history, evaluation was considered to be more of a separate activity, rather than the more integrated model between evaluations and programs that exists today. This original plan, however, proved to be challenging for the Foundation for several reasons. The same program officer already quoted sheds light on this as s/he explains that by the time results from these evaluations came in, it was too late for them to be actionable, and the Foundation's priorities and objectives could change in the meantime.

One program officer explains the shift in greater detail, "I want to start planning that evaluation at the beginning as opposed to doing it four years into it and trying to back build, which doesn't seem to work very well." Even though it was decided several years ago to shift the timing of evaluations so that they started earlier, the process of normalizing it among staff is ongoing. This is exemplified as one officer shares, "We've been trying to work and get everybody to speed with the idea that you start thinking about the evaluation at the same time that you develop the program, which is the beginning."

Even though Foundation staff agree that this change is ultimately for the better, there are still some challenges, particularly considering that programs evolve and change in the early part of their existence. One officer explains this and shares how the Foundation is working to address it, saying, "It is very difficult for the evaluator, if then a year later for a program, we just decide to do it this way instead of that way. So I think we're trying to deal with that, see what kind of flexibility we can build in, ... We make sure at the very beginning, [that] we have the core elements right, and then try not to change them." This shift also works to make evaluators more like collaborators with the program. As an example, an officer shares, "So once [evaluators] went in earlier and they discover something, shouldn't they share that with the program that we're trying to make better? But then the evaluator's part of the program, but you know what? That's okay."

The desire for more useful and actionable results ultimately motivated this shift in policy. Now the Foundation is better positioned to use evaluation results for program improvement purposes, or to inform its currently active strategic objectives. There might be a trade-off, however. One member of the Foundation leadership shares a concern, “What I’ve been worried about lately is whether we are evaluating too soon and whether we should, I mean, not that we shouldn’t evaluate soon, but maybe we want to let the thing go on a little long to see.” The RWJF leadership member raised this concern because in some cases, the programs funded by the Foundation are trying to impact effects that are very distal to the intervention. In the rush to produce evaluation findings quickly, the Foundation might be short-changing itself on learning about some of the longer-term outcomes.

Ultimately, both these policies changed because the thinking that justified the old policies changed. In the case of lack of evaluator collaboration, the Foundation’s beliefs shifted away from thinking that model was the best way to get credible results. Regarding starting evaluations earlier, the Foundation’s evidence interests changed. Initially, the Foundation was most interested in getting information about whether or not a program worked. Now, in addition to that, the Foundation is also interested in more timely findings that may be used for other purposes (like informing program improvement). In both of these cases, it seems that the origin of a policy change is actually a shift in Foundation values and beliefs. The policies then slowly evolve in informal ways to mirror the new values.

One of the research and evaluation officers in a leadership position was able to offer a few more explicit examples of how s/he was able to bring about a shift in policy. The first had to do with a desire to ensure that all evaluations included the development of a program logic model. The officer introduced the concept through sessions among the staff. Ultimately, this has become a common practice because staff members see the value in it, not because it is a hard and fast policy.

The second example is this same officer's introduction of evaluability assessments. In this case, the officer used examples from practice to share how the practice was beneficial; s/he shares, "Most people can't really envision what this is supposed look like unless you tell them a story. As soon as I told the story of evaluability assessments, other people could see, 'huh, I could see where I could use that with my work.'" This process of introducing staff members to new policies through example and relevant practice examples is still challenging. The officer shares, "At this point I've got some precautionary tales for the staff and some really splendid examples of how it can be when we do work with clarity with the grantees. And I try to use both to continue to exhort this kind of change. I'm not always successful. In fact, it's extraordinarily imperfect."

CHAPTER 5

THE IMPLEMENTATION OF EVALUATION POLICIES

Introduction

The implementation of policies is best viewed through the case studies. This chapter presents the cross-case synthesis and analysis demonstrating how the evaluation policies as described through the documents and interviews are implemented in field settings. The discussion loosely parallels the broad categories of evaluation policies in the previous chapter, but it is not an exact reflection. The glimpse into implementation of the policies begins once the evaluation has started. It is a look mostly at the operational policies. As such, this part of the chapter does not include assumptions that underlie the policies. Rather, the discussion begins with how the evaluation policy goal of use manifests in the implementation. Then, examples of how policies pertaining to decisions about what and when to evaluate are implemented are presented. This is followed by information relating to how evaluators are selected and contracted, and how evaluations are designed and conducted in the field. The implementation of policies related to administration of evaluations and stakeholder involvement are then discussed, followed by how findings are reported in practical settings. Finally, information about the conditions that support the successful implementation and barriers that impede implementation are discussed.

The Interpretation of Evaluation Policies

An important finding in this study is that there is a significant difference between how policy goal statements and policies that are operationalizations are treated when it comes to implementation. As previously discussed, there is a high degree of consistency when it comes to evaluation policy goal statements. This is confirmed through an examination of the data presented by the case studies. However, as was suggested in the previous chapter, there is less alignment

concerning operational policies. Individual program officers implement operationalized evaluation policies on a case-by-case basis. While this is never stated explicitly, it seems to be a tacit understanding of how the work is to be done, and a function of implementation of evaluation policies that are largely informal and unwritten.

Despite the inconsistencies, in some ways, it seems as if the operational policies are being interpreted as intended. The operational policies are oftentimes written in a way that might promote flexibility—they serve more as guidelines to suggest how practice could be done. The Foundation seems to rely on the expertise of people it hires and contracts to be the main determinants for how the policies are enacted. One program officer summarizes this nicely, saying, “I don’t think we really have an evaluation policy. I think we evaluate a lot of different things. I think the policies to try to use the best design for the situation. I really do. I think depending on who you are, and what the situation is, you as a staff member might say, ‘I really think this is the strongest design for the situation.’ In general, I think the people here are really good.”

Allowing particular program officers to determine these operational aspects of the evaluations based on their experiences and the setting affords a certain amount of flexibility and adaptability when it comes to evaluation practice. While responsiveness to context is clearly a benefit to having policies that are not over-codified, one consequence is that there are inconsistencies across people and projects about how evaluation is practiced. These issues are explored more in depth as particular interpretations come to light through careful examination of the case studies.

Implementation of Evaluation Policy Goals

The evaluation policy goals reflected in the case studies show the emphasis that the Foundation places on use of findings. In interviews on each of the cases, different types of use of evaluation findings are discussed. These examinations also shed light on how the desired use of findings can in many ways shape evaluation design issues. The specific Foundation policy goals that

are echoed in this part of the analysis include the desire to build an evidence base around a strategy and inform program improvements.

Because the Holistic Health program and evaluation (the complex case) represents such a significant investment of Foundation assets, there is a lot of emphasis on using findings from this evaluation to inform the field. Both the evaluator and the Foundation have a clear interest in publishing findings from this evaluation in peer-reviewed journals. One Foundation staff member shares, “The evaluation findings are reported in the scientific literature because we wanted that, and if we’re doing such a rigorous design, it probably warrants that. But it usually doesn’t get traction unless it’s somehow citable in the professional literature.”

The last part of this quote is particularly interesting. According to this Foundation employee, ideas about interventions and programs are most effectively spread through academic literature. This desire for a particular use of findings ultimately ends up shaping how the evaluations are designed. This will be further explored in a later section.

Another example of using evaluation findings to contribute to an evidence base is found in the Drug Free Families program. This proved a more challenging example, however, as it was a chief point of contention between the evaluator and the program director. According to the program director, the evaluator’s mindset was too academic to the point where the design sacrificed the ability to capture more nuanced findings that would suggest positive program impact. Ultimately, the program director felt the Foundation’s and evaluator’s interests in using findings conflicted with his/her own intentions for using the findings in service of influencing federal policy and appropriations. Ultimately, even though the findings were not positive, the Foundation was able to use them to inform its grant-making efforts. As one Foundation leadership member said, “The bottom line lessons here are that bringing about change like this program wanted to bring about is

very difficult to do. It's a negative set of findings as opposed to a positive, and negative findings are often just as useful to avoid spending money on things that aren't going to have an impact."

Using findings to support program improvement was also seen in several of the case studies. This emerges most prominently in the Fun and Games case. Findings from the initial implementation study were used in a variety of ways to inform both specific improvements in program delivery and the strategy used to scale up the program. Specifically, through the implementation study, the program learned about more efficacious training designs, and the ways in which contextual factors influence program implementation (such that those factors could be better addressed as the program went to scale). Part of the reason that this model was so successful in this case was that the program leadership was very receptive to the information. As the evaluator shared, "[The program] didn't want to hear that everything is great and they're doing an awesome job, but they really wanted to hear what was working. [The program was] trying to improve, and they were going to be expanding and they wanted to know."

Even though the pilot and implementation study phase of the Holistic Health program was truncated, there were still opportunities to use evaluation findings to support program improvements. The evaluator explained, "[The program has] learned from our work, and they made programmatic changes in terms of requirements and technical assistance ... related to that. I think our work has contributed to their thinking and the program changed as a result."

Aside from program improvement, the Fun and Games program was able to use evaluation findings in additional ways. Specifically, the program used results from both evaluations (the implementation and impact studies) to aid its marketing, funding, and expansion efforts. The evaluator explains, "the impact of the evaluation is several fold; one, obviously is simply the evidence that promotes Fun and Games, but it's also the communications platform that it provides. It's the funding that it attracts from funders who only fund evidence-based programming. It is the

marketing ability it gives as [the program] go[es] out into different sites. There is a lot of collateral impact from that evaluation.” These ideas were confirmed in the interviews with the program personnel.

As discussed in the previous chapter, the evaluation policy goals heavily influence the operational policies. This same pattern plays out considering the implementation of those policies. This will be explored in the next several sections of this chapter.

Deciding What and When to Evaluate in Practice

These case studies offer examples indicating how evaluations are scheduled. As noted in the previous chapter, it is a Foundation policy to begin evaluations in concert with the start of program activities. Also, this is a policy that has changed over time at the Foundation. The chronology of this shift in policies matches what happened across these three cases. The oldest program, Drug Free Families, is the one example in which the evaluation was commissioned after the program had begun. In the other two instances, the evaluation either started with program activities, or it preceded them.

The other concept that emerged in the previous analysis is the idea that evaluations should match the stage of the program. In the case studies, this worked very well in the Fun and Games program. An initial implementation study was commissioned that helped inform the program as it moved into the demonstration phase. As this occurred, an impact study was commissioned. The Foundation took measures (specifically requiring that the implementation evaluators be included as partners in the impact study) to ensure a smooth transition across the two evaluation efforts. The implementation study helped to inform some of the outcome measures investigated in the impact study.

In the case of the Holistic Health evaluation, the intention to have a pilot phase was not followed as anticipated. The program scaled up much faster than originally planned, which put the

evaluator in a situation in which the evaluation had to be modified significantly after it began in order to account for the programmatic changes. Even though this worked out in the end, and everyone is very satisfied with the process, it was not as smooth as the Fun and Games example.

In the Drug Free Families case, there was an implementation study commissioned in the initial pilot phase, but it was completely independent from the impact study. In fact, results had very little bearing on the program scaling efforts, or the impact evaluation that followed. This did technically follow policy for the most part, but it did not play out in a successful way in the field.

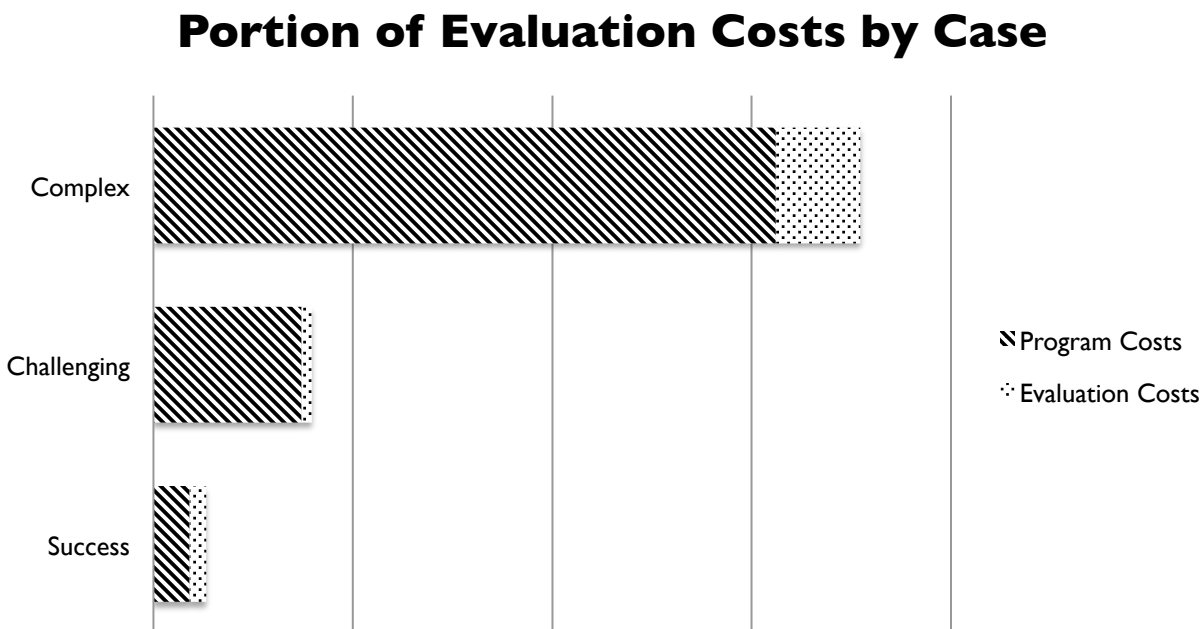
Implementation of Evaluation Policies Regarding Selecting and Contracting Evaluators

The document analysis and initial interviews provided information about some of the logistical aspects of evaluators (including contracting and resources) and the selection of evaluators. The case studies provide specific examples of budgeting for evaluations and the processes involved in selecting evaluators.

Evaluation resources. In the first phase of this research, the Foundation documents suggest that 20% of its overall spending is devoted to research and evaluation efforts, but the division of monies between these two activities is unclear. Examining the case study evaluation budgets as a percentage of the overall budget sheds some light on how evaluation budgets exist in practice. These three cases represent this wide range of funding. In the successful case, Fun and Games, the evaluation costs represented 6% of the overall budget. In the challenging case, Drug Free Families, 32% of total costs went to the evaluation, and in the large and complex case, Holistic Health, evaluation costs made up 12% of the overall budget. However, these percentages only tell part of the story. Each of these programs was funded at a very different level overall. To avoid sharing the exact amount of grant awards (which may facilitate specific identification of the programs), a comparison without numbers is presented in Figure 4. This visual representation offers

a relational sense of how much money was devoted to funding the evaluation efforts and program activities of each initiative.

Figure 4
Portion of Evaluation Costs Compared to Overall Budget by Case



There are significantly more resources devoted to the complex case than both the successful and challenging cases. In fact, nearly four-and-a-half times more resources are expended on the complex case than the challenging case, and nearly 13 times more resources than on the success case. In calculating the overall ratio of evaluation costs to program costs, across these three programs, evaluation efforts represent a little over 12% of the budget. Granted, this is a very small sample, and this information should not be used to draw inferences Foundation-wide.

Evaluator selection. So far the data has shown that evaluators are selected and contracted through two primary methods: sole-sourcing, or invited bid. The case studies provide examples of how this works in the field, and the implications of those methods.

The success case had two separate evaluators contracted. The first, for the implementation study, was sole-sourced by the R&E program officer. The team was selected because of its extensive content expertise and experience evaluating programs in similar settings. The second evaluation, the impact study, was contracted through an invited bid. The first step in this process was developing a request for proposals, led by the R&E program officer. It represented a collaborative effort between the R&E officer, the evaluator from the implementation study and program personnel. The process is explained in the following quote from one of the program staff, “We worked very closely with [the R&E Officer] to develop the request for proposals from the Foundation to the evaluation firms. And the Foundation definitely drove that, which was good because I don’t have a background in evaluation. The Foundation does, so all of the questions that we needed to ask, it was great about. We had input into the questions that were being asked and the description of the study’s purpose. We had a lot of input into that.” This demonstrates what was conveyed earlier in the analysis—that frequently evaluation questions originate with the Foundation. Also, it provides an example of collaboration between parties on evaluation efforts.

The Foundation chose to interview two of the firms that submitted the proposal. Again, the program personnel and implementation evaluators were involved in this process. In fact, one of the requirements for the impact study was that the selected evaluators would have to work with the research center that conducted the implementation study. One of those team members explains, “Because we were so familiar with the program already, [the Foundation] wanted to keep us in there.” This also played into the eventual selection. According to the R&E officer, “Part of our selection was based on prior familiarity of the [research center that conducted the implementation evaluation] with personnel at those other two places. In other words, that [the implementation and impact evaluators] would get along, they would be comfortable immediately so you wouldn’t have all kinds of problems.” Interviews were conducted over the phone and featured all of the previously

mentioned parties. Ultimately, the evaluator that presented the stronger design in terms of the methodology was selected. According to the lead evaluator on the project, there was some back-and-forth about the proposal to clarify questions that the Foundation had. Once the evaluator was selected, a précis was completed by the Foundation and the evaluator that includes information about the scope of work, evaluation activities, project deliverables, and budget. This is what the Foundation uses as its contract.

The Holistic Health evaluation represents an example of a sole-sourced evaluation. The evaluator, was contacted, as s/he explains, “because of my work on health policy. I was told that some of the work I did topically, people looked at. [The Foundation] looked up files about some of my work and, because it was relevant, that’s how I got in contact with the Foundation.” The evaluator was involved with Holistic Health for a significant period of time before s/he was officially designated as the evaluator. Specifically, when the Foundation was doing some advance work preparing to introduce the program, the evaluator was informally consulted for some opinions about the program design. At one point in this process, s/he was asked to write an evaluation proposal. The evaluator was asked to interview with the Foundation because there was an interest in assessing his/her flexibility, and “how comfortable [the evaluator and the team s/he had selected] were [with] being really enmeshed with the program, as opposed to being really objective outside evaluators.” Over the life of the program and the evaluation, there were three separate contracts. Each was followed with a précis following the same basic format as described above.

The impact study of the Drug Free Families program presents an interesting case because there was a very rigorous process involved to select the evaluator that was overseen by the Evaluation Advisory Panel. According to one Foundation leadership member, “Everything about this program was elaborate. The evaluation process was elaborate. You know, it was a big investment and it was something that everybody cared about.” According to the R&E program

officer, “We certainly had a competition to pick the evaluator. We had a solicitation [to between 10 and 12 firms]. There were only two submissions that rose to the level of funding, and we had both teams come in and present to us.”

The process was described as both formal and participatory. As explained by one Foundation staff member, “It was a very formal process. This was a very blue ribbon evaluation, with an advisory committee. And the advisory committee really was an active partner in making the choices about design issues and about selection of the contractor or the grantee.” In addition to the participation of the advisory committee, the program director was also very involved in the selection of the evaluator. The process was described as being, “very participatory, where we involved the program side a great deal in designing and choosing the evaluator.” The program director agreed with the final selection because s/he felt that the firm that was not selected offered a design that was too simple and not attuned to the details and complexities involved in the program.

This is interesting because, of the three cases, this was the most rigorous and elaborate selection process with the greatest number of proposals solicited. This most closely matches what was earlier described as the preferred method for contracting an evaluator. However, this evaluation ended up being the most problematic of the three investigated in this study. The program director even attributes many of the challenges faced in this process to having made the wrong choice in selecting this evaluator. Furthermore, the greatest proportion of resources were dedicated to this evaluation in comparison to the other two. The evaluation with the smallest relational budget was the success case. Again, generalizing based on only three instances is not possible, but this does suggest that evaluation funding levels might not be an important determinant of evaluation quality.

Implementation of Evaluation Policies to Design and Conduct Evaluations

The case studies provide valuable insight into the implementation of policies governing evaluation design and activities. The information that emerges through this part of the investigation

focuses mainly on the process of designing the evaluation and data collection and analysis activities. Several overall trends emerge. First, in the initial phase of research, understanding the program was treated as a distinct activity from designing the evaluation. However, as this plays out in real-world settings, these activities are enmeshed with one another. Second, the Foundation favors collaboration over independence when it comes to designing the evaluation, but independence of function is more strictly adhered to during data collection and analysis. Finally, external factors influence and shape evaluation designs.

Design. The design of the evaluation for the Drug Free Families program was challenging from the outset, mainly because it was very difficult to determine and operationalize dependent variables. Part of this challenge lies in the fact that the ultimate desired outcomes (less substance abuse) would not be able to be ascertained until more than a decade after the intervention. As one member of the Foundation leadership explains, “That’s sort of an evaluator’s nightmare: the ultimate effects are in flux.” Ultimately, the evaluation sought to measure intervening indicators and the family and community level, “the idea was to work with families and communities, strengthen them, and then that would lead to better raising of the children, better environments for the kids to grow up in.” The evaluator indicated that the process for determining dependent variables to measure involved several parties. s/he explains, “We worked with the program office and our evaluation advisory panel to identify ... indicators of each of those that the program was intended to, and might be likely to affect.”

The most significant evaluation of the Drug Free Families program was an impact study that featured measures of both process and impact. According to the evaluator, “We wanted to try to have the strongest and most rigorous design that would be realistic. We couldn’t do random assignment, but we wanted to have the strongest design in order to be able to have fairly strong conclusions at the end.” This desire informed the choice of a quasi-experimental design in which

matched communities were selected to facilitate comparisons between communities that received the intervention and similar communities that did not. The evaluator submitted ideas about a design in the response to the RFP, and shares, “I think the design that we proposed ended up getting changed pretty substantially. You know, [the Foundation] like[s] our design, and us as a group well enough to choose us to do the work, but there was a lot of evolution of it before the design was finalized.”

Several people were involved in this process of finalizing the design, including Foundation program and R&E officers, the chair of the Evaluation Advisory Panel (EAP), other EAP members, the evaluator from the implementation study and the impact evaluators. The process of designing the evaluation was described as a “negotiation.” This was explained further by a R&E officer, “A lot of early time with the evaluators was spent identifying what will be our dependent measures, what will be our independent measures; how will we control for community differences?” The program director, also a “capable and prominent” researcher had “strong opinions about how it would get designed” that were taken into account.

The EAP was also very actively involved in the process. The evaluator shared, “We had a very engaged and active evaluation advisory panel that the Foundation had put together. And so [the advisory panel] was quite influential. It’s not like [the advisory panel] always spoke with one voice, but we listened very carefully to that because it really was quite a stellar panel with very impressive expertise and experience, and so we’d listen to them.” Ultimately, even though the results from the evaluation were disappointing, everyone except the program director considered the design to be a strong one (in that it met commonly accepted standards of rigor). The program director, however, disagreed, and felt that the design was not sensitive enough to detect all the changes that were happening in the communities.

The Fun and Games program was initially evaluated through an implementation study that was described as “pretty open-ended” and “qualitative.” The questions guiding the work were “How has the program been implemented? What are the contextual factors that affect the quality of implementation? How is the program received?” The initial study took place over one school year and focused on eight schools in one town, six of which were newly implementing the program, and two that had a lengthier experience with implementing the program. This design was reached through collaboration mostly between the evaluators and the program. The evaluator explains, “We developed the design and the methods, but [the program] definitely had a voice in that because we wanted to know what [the program] wanted to know. If your work is going to be useful, you have to find out from the partners what will be helpful to them.”

The information gathered through this implementation study played a large role in influencing the design for the impact study, which maintained a focus on evaluating the program processes, and added aspects of evaluating outcomes. The implementation study was particularly informative in helping to determine what the appropriate outcome measures should be. The evaluators for the implementation study also remained as partners in the impact study (a requirement specifically written into the RFP). According to the program director, “the goals were to document the outcomes using a randomized design so that we could communicate to the world about our impact.”

Prior to sending out an RFP, the Foundation knew that it wanted to investigate the effects of the program with a randomized control trial (RCT). This preference is due in part to a belief held by many at the Foundation (and in the field in general) that this type of design is the most credible way to determine an unbiased estimate of program effects. The Foundation had indications through the implementation study that this was a successful model, and wanted to be well-positioned to share more rigorous findings with a very wide audience. One of the program officers explains, “as

we scaled, we knew that we could get to the 27 cities flagship model, but to go further than that, one would need a randomized control trial to make the case that this should be universal; to make the case that this should be policy.”

With this preferred method in mind, the implementation evaluator worked with the Foundation R&E officer to help develop the specific RFP and the outcome measures of interest. Determining the outcome indicators involved operationalizations of some broad and hard-to-measure constructs. This was aided by the data gathered through the qualitative study of the program implementation. The impact evaluator explains, “Through that (implementation) study, [the Foundation] had gotten kind of anecdotal information about some areas where the teachers thought that things were improving. And so I think that did help us focus our data collection efforts on this current study.”

Once the evaluator was selected, the impact study design was finalized. In addition to the Foundation and implementation evaluator, the program personnel were involved in this process. According to the program director, “We really informed the design ... a great deal. I met with the evaluators in person and we did the school recruitment [into the study] because we needed it to match the way we did school recruitment normally.” The evaluator went on to say, “the program themselves actually were involved in helping us [design the evaluation] as we started to think about what outcomes to look at, and what scales to use for various domains on the surveys. We did have multiple discussions with the program to make sure they thought we were capturing what they as a program intended to affect.” When asked about this process, the Foundation officer shares, “I think that’s a great lesson learned for how to think about structuring evaluations—alongside the strategy for developing and growing the program. So it’s not an evaluation coming in after the fact; it’s not an evaluation designing the program; it’s an evaluation working alongside of the program strategy being developed.”

Aside from bringing in the various expertise and knowledge of the people involved in the evaluation, the design was influenced by several external factors. First, there were structures in place that supported the successful implementation of the RCT design. The way that the program was rolling out facilitated the randomized control trial. The number of schools interested in having the program exceeded the available space, so the national program office maintained a waiting list. Schools were matched to one another prior to randomization. Schools that had to wait a year received a discount on services the following year. The design was further influenced by the desire to be able to publish results in a national database for rigorous research studies. Initially, the evaluation was budgeted for 20 schools, but in order to qualify for the database, research activities needed to take place in 32 schools. After a presentation to the Foundation staff, the R&E program officer was authorized to increase the evaluation budget to be able to reach the higher number of schools.

Another factor (external to the program) that influenced the evaluation design was a desire for findings to be able to speak to one of the Foundation's key strategic objectives: childhood obesity. This was not an initial goal of the program—their interests lie more in social/emotional outcomes (like better conflict resolution) and academic gains (faster transition times to being on task), than increases in physical activity. The implementation evaluator explains, “because the Foundation was the key funder, it overlaid this physical activity piece on it. The program itself was funded under a specific strategic objective wherein the physical activity goals were more important, and so [the Foundation] overlaid a piece of it that was about physical activity that we had never heard before. And that Fun and Games wasn't really talking about.” Physical activity levels were thus added to the design at the urging of the Foundation.

The evaluation of the Holistic Health program utilized a quasi-experimental design (often employed when randomization is not possible or appropriate) to reduce bias in inferences about the

strength of the program. According to the R&E program officer, the quasi-experimental design was a critical part of the evaluation. Even though it was fraught with challenges as both treatment and comparison community contexts changed around the intervention, everyone was united in their interest in having a defensible (according to the academic world) design. Instituting this type of design in a real-world setting is always complicated. One of the specific challenges in this case was that “the initiative was still evolving.” In commenting about the design, the evaluator explains, “It’s trying to make academic and scientific sense of the messy real-world realistic evaluation. An ongoing challenge is that this isn’t an experiment. Nothing is randomized here. So were trying to kind of learn based on what’s happening, and doing it in a way that is scientifically credible and will meet the merits of peer review.”

As in the other two cases, the process of designing the evaluation was very collaborative, and featured input from the Foundation staff, program office, and evaluators. Despite this characterization, the evaluator shared, “I think we were smart enough to realize that [the Foundation] was paying for this and, at the end of the day, we need to kind of get it right. We need to do something that’s interesting to them. And I think we approached it as a negotiation probably more so one our end than on [the Foundation’s] end.” Even though the evaluator had that mindset, s/he still felt confident to suggest changes, and shared an example, “I think in some cases, there were some conversations where we might have had to do some convincing that you know what, you guys may not think this is important, but we really do, based on our background and experience and, as such, we think that this needs to be an area that you should focus on.”

Because the evaluator was brought in very early on the project, s/he was able to be a part of the early discussions about what the program was supposed to accomplish. The evaluator was a major player in the development of the program logic model. This early participation helped pave the way for helping the evaluator adapt to changing circumstances. The initial evaluation was

designed to collect data to inform the Foundation about project implementation, and to suggest data that might be gathered in a later summative evaluation. This, however, changed when the Foundation decided to scale-up the project in under a year. The evaluator shares his experiences, saying, “Initially what changed for us was the scale, because the Foundation changed its thinking. And also we were thinking that this would be more of a traditional pilot—we learn from it, [the Foundation] would learn from it, and we present our results and then [the Foundation] would decide if it wanted to scale up or expand or change or modify. And very quickly [the program] went from 4 communities to 14 communities [and the project eventually settled at 16 communities].”

The design here was also informed by external factors. Specifically, the evaluator had his/her own goals about what s/he wanted to get from the evaluation, explaining, “I’m an academic. My interest in taking this on is not just to get paid to do something for 10 years, but it is to inform the field around the theory that communities could make a difference ... The goal was to design a study that would have scientific merit and allow us to publish.” The desire to inform the field weighs heavily on the types of designs that are permissible for this type of intervention.

Even though the gold standards of research (using RCTs and quasi-experimental designs to reduce bias in inferences) prevail at the Foundation and represent a common goal, one member of the leadership expressed some concern about the value of this kind of design in this type of long-ranging and large project that is deeply embedded with a constantly-changing context. Specifically, the leadership members shares, “We also have been following a set of control communities in there. It’s a little bit of a challenge. The environment that existed in 2006 is not the environment that exists now. Part of this is that the kind of change we’ve been working on in these communities has ... not been pristinely isolated from the rest of the world ... So essentially what were going to be concluding is [through] measuring the rate of change compared to the control communities. This is not going to tell us the effectiveness of the program.”

Conducting evaluations. These cases also provide field examples of how data collection and analysis policies are implemented. The three projects show a range of quantitative and qualitative approaches to data collection and analysis. These include interviews, surveys, focus groups, observations, and secondary analysis of large national data sets and other available public records. What emerged through the cases echoes what was indicated in the policies: the Foundation is heavily involved during the design phase, but is much more hands-off during data collection and analysis.

An aspect of evaluation practice described in the first phase of the study for which corroborating evidence was not gathered is activities relating to developing conclusions based on the findings.

Implementation of Evaluation Administration Policies

In the first phase of the research, the administration domain includes information about various roles assumed by the different parties during an evaluation, and the processes (like communication and managing relationships) that help support these roles. The information collected in the case studies does not map directly on to these categories. However, evidence of collaboration and communication across the projects sheds light on how some of these administrative tasks play out in the field. The specific areas that emerge through the case study analysis here are collaboration, ongoing communication, and resolving conflicts.

Collaboration. As has been previously mentioned, the Foundation currently takes a very collaborative approach when conducting evaluations. These three cases offer examples where this worked very well in serving the aims of the program, and when it proved to be challenging. The success case and the complicated case both were very collaborative. In the complex case, a Foundation leadership member describes the collaboration as “a constant engagement of people at the Foundation, between the people who are managing the program and the people who are

evaluating it.” The Foundation leadership member further went on to describe how at the time, this was not the norm for relationships on evaluation projects funded by the Foundation; s/he says, “So I think we learned a lot about how to do that, I don’t think that most evaluations of the Foundation were like that when we started and I’m very proud of that.” According to the R&E officer on this case, the collaboration here was born partly out of necessity; s/he offers, “you have to get to know your evaluator well, especially for something like this that goes on and on and on. The evaluator became like a partner in the work and was viewed as a partner.” The collaboration here helped facilitate the evaluator’s ability to adapt to the changing context and demands of the program. Ultimately, the evaluation was flexible and responsive in part due to the collaboration between him/her and at the Foundation.

The success case also benefitted greatly from collaboration, largely because of the relationships between the implementation and impact evaluations, and the ongoing participation of the program director in both evaluations. Because the program director and evaluators worked closely together, the evaluation design of the impact study was able to meet varying needs of the Foundation and program. It also was informed by the work previously done. This collaboration was partly mandated by the Foundation, as it was required in the evaluation contract. However, the parties did not collaborate begrudgingly; their relationships were characterized by a great deal of respect for one another and common goals as they related to the program and its evaluation.

Efforts for true collegial collaboration were not met with success in the challenging case. The stage was set to support more collaboration—the program director was directly involved in helping to design the evaluation and in selecting the evaluator—but challenges proved too much for the relationships. The collaboration fell apart as results from the evaluation were disappointing to both the Foundation and the program.

Ongoing communication. Collaboration efforts are generally supported by ongoing communication between parties. In the cases, it seems that this is generally brokered by the Foundation. All of the evaluations had some sort of pre-established plan about how and when the various parties would communicate about the evaluation process and progress. In all cases, it was reported that there is generally more contact in the beginning, when evaluations are being designed, and at the end as a reporting strategy is needed for conveying findings. In the interim, when data is being collected and analyzed, there is less contact between the Foundation and the evaluators. This directly reflects what was discussed in the first phase of the analysis regarding the relationship between the extent of communication and the phase of the evaluation.

More specifically, the Foundation program officers generally set up a system by which there were regularly scheduled calls and/or meetings. At the times of heavy collaboration, this might have been every week or every other week. When communication was at its lightest, these calls took place every month. Also, even though they were regularly scheduled, calls were subject to cancellation if the parties felt as if there was not enough new information to share. Aside from the regularly scheduled calls, Foundation officers indicate that they were always available by phone and email to the evaluators. Even though this seems standard practice for the Foundation, some of the evaluators indicated that it was unusual based on their previous experiences in other contexts. One notes, “At first that seemed a little odd to me. I was used to the sort of prior grants where you get the grant, and then you’d go do your thing. I was worried about our independence, since we were contracted, and I didn’t want anybody sort of being able to influence that at the Foundation. So it seems a little bit odd, but after going through it, it was critical because we would need information on program changes. There was a lot of back-and-forth, but in a very collegial way and not in any kind of threatening way.” This speaks to the fact that this evaluator was accustomed to a more independent

model of evaluation; however, the Foundation's focus on collaboration proved to be efficacious in this case.

The benefits of this model of communication were also noted by one of the program directors, who explains, "We included our R&E officer as much as we could in every conversation, particularly the ones where things were challenging because s/he knew what we were trying to do so well, and was willing to advocate for us. And then, [the R&E program officer] also helped me learn about the limitations or the priorities of evaluation that I didn't understand before." The regular involvement and constant communication of the R&E officer was beneficial to both the evaluation team and the program in this case.

In general, the content of these calls, as described by one evaluator, include, "provid(ing) updates on progress. The Foundation would ask questions about how things were going, so we were meeting once a month with them. Then, we would give updates on what we'd accomplished over the past month and talk about any issues and proposals for how to deal with the issues."

In the challenging case, the points of contact increased in response to some of the difficulties experienced. The evaluator recalls, "at that point, we started having very regular calls, which involved the Foundation, the chair of the advisory panel, the program director and deputy director and the eval(uation) team. So we started having very regular contact phone calls and periodic meetings."

Resolving conflict. Finally, the case studies offer insight into what happens when projects encounter specific challenges, or there are conflicts between parties. There were instances in all three cases of some sort of difficulty that arose during the evaluation. These ranged from minor instances in which signals got crossed, or people were left out of a communication loop to major external circumstances that derailed or otherwise severely impacted evaluation activities. These challenges were addressed in different ways across cases. Some examples are shared below.

In some cases, evaluation activities and communication became challenging as Foundation personnel either left the Foundation, or were assigned to new projects. This resulted in challenges for the evaluators to bring someone new up to speed on the project, or to manage different expectations regarding the projects. The involved evaluators mention that this is a minor distraction, but one over which they were powerless. One R&E program officer mentioned that in this case, there was a transition memo to assist with the change, but it is unclear if that served to alleviate any of the difficulties experienced by the evaluators.

Another instance of challenges relates to specific communication misses or missteps. In one instance, the evaluator introduced a new measure without first sharing it with the program. The program personnel became upset. To resolve this, the R&E officer called a meeting between all parties to discuss the issue. This seems to have been an idea that originated with the program officer; it does not seem to be reflective of a larger policy. Another confusing communication instance arose when Foundation reporting channels were not clear. In one of the cases, the program was funded under one team, but the evaluation was funded under another. This relates to the first example in which there was movement among program officers either away from the Foundation or to a new team. Occasionally, the evaluator would communicate with one of the teams, but not the other. Eventually, communication became problematic enough to warrant the development of a communication agreement. After which, things ran smoother. Again, this seems to be an action born out of particular circumstances, rather than as a result from an overarching rule or policy.

Finally, there was an issue in which the circumstances of the evaluation became incredibly difficult, and they manifested in tensions in the relationships between the people involved. This happened during the challenging case. In essence, the impact evaluation was unable to demonstrate that the program was effective in reaching its goals. This upset the program director, partly because s/he was very personally and professionally invested in the program. A Foundation staff member

explains, “The program side often starts with a strong premise that what they’re doing is good, and it’s very hard when it doesn’t work out. I saw that time and time again, and it’s not because the program is bad. It’s often because the problem is so much bigger than the intervention. It’s very frustrating because the intervention might be a good one, it might be a necessary or possible condition to bring about change, but then you have all these environmental factors and all these other issues that get in the way, and it’s very hard to be told that your intervention is not working.” Relationships became very strained, and the Foundation ended up playing, according to the evaluator, “a pretty critical role in sort of facilitating and brokering those conversations.”

Implementation of Evaluation Policies Pertaining to Stakeholders

The significant emphasis on stakeholder involvement in the Foundation records is not mirrored in the case studies. Interviewees across the cases were asked to identify whom they consider to be key stakeholders in the evaluation, and to describe the ways in which the stakeholders were involved in the evaluation process. The involvement of stakeholders in each case is described below.

In the Holistic Health program, the R&E program officers identified the Foundation, the program communities, communities doing similar work, researchers in health systems, federal policy-makers, and those trying to improve the quality of health systems as stakeholders. The evaluator largely echoed this list—s/he mentioned the Foundation, followed by policy-makers, practitioners, and academics engaging in similar research.

These stakeholders were engaged in a variety of ways in the study. The relationships between the evaluator and the intervention communities evolved naturally, over time, as there was frequent and close contact between the evaluation team and the communities over an extended period. The evaluator explained the various ways in which the communities were engaged in the study, specifically through, “sharing our research findings, and meeting requests that [the communities]

have for information.” The evaluation team also worked with stakeholders to share findings in speeches or at board meetings, and to answer informal questions that the stakeholders may have had. Finally, the evaluator shared that they consulted stakeholders to get their “feedback on different aspects of the evaluation and its design.” In conclusion, the evaluator offered, “So we’ve done, I think, a really good job, and we’ve had very good relationships with stakeholders in that sense.”

Beyond the relationships with the communities, the R&E program offer indicated that the evaluator also had good relationships with the Foundation staff and the national program office. The R&E officer further explained that developing these relationships was critical because it provide means for collecting data. These relationships were in some ways brokered by the R&E program officer; s/he says, “I worked really closely with the evaluator and his folks, and tried very hard to facilitate direct contact between the program folks in the evaluator.”

In the Fun and Games Case, the two different evaluators identified similar lists of stakeholders including the program, the Foundation, schools, and policy makers. However, there was a difference in how these lists were prioritized. In the implementation study, the program was considered the primary stakeholder, whereas in the impact study, the Foundation was identified as the primary stakeholder.

Relationships were built between groups early on in the program with the assistance of the R&E officer; s/he shares, “At a very early date, we got the impact evaluation firm, the center that conducted the implementation study, and the program folks together to make a concerted effort to develop logistics together.” The program director explained his/her interactions during the evaluation, saying, “I personally dealt mostly with [the evaluators] to solve problems around the pieces that we were responsible for, or to fill in the context when [the evaluators] were wondering about why something was different at a particular school or something like that. So I interacted a fair amount with [the evaluators].” Beyond the program and Foundation, the nature of relationships and

extent of involvement with other stakeholder groups was less clear. In some cases, school administrators were relied on for access to data, and district-level reports were provided when interest warranted them. These relationships seemed to unfold on a more as-needed basis, rather than emerge as a part of a comprehensive stakeholder engagement strategy.

In the Drug Free Families program, many of the same broad categories of stakeholders were named (the Foundation, the program, and the related practitioner and research communities). There was more of an explicit and direct focus on the federal government as a key stakeholder. This is because the program was trying to influence and shape public policy. During the evaluation, the evaluators had meetings with the relevant federal government members.

One of the chief challenges in this evaluation was the relationship between the program director and the evaluators. Despite being directly involved in the selection of the evaluation team, the program director eventually felt that the evaluation was not working to meet his/her needs. S/he had very specific goals in mind for what s/he wanted the evaluation to accomplish related to helping advocate for the program to the federal government. Despite oversight and involvement from an Evaluation Advisory Panel that worked to mediate some of the conflict, and the evaluators' efforts to "try to work with [the program], and be as responsive to its concerns and thoughts and needs as [they] could be," the relationship issues were never resolved.

What these three cases show is that stakeholder relationships manifest in a variety of different ways in field settings. Based on these findings and what emerged during the first phase of interviews, it seems as if stakeholder involvement is determined more on a case-by-case basis based on the specific program needs and context at hand. The evidence does not show use of types of strategies specifically recommended by the Foundation to support stakeholder involvement. The involvement was more indicative of a practice that emerged from the situation, rather than a top-down approach. If the Foundation was interested in evaluations taking a more directed approach to

involving stakeholders, one way to ensure it would be to write more specific strategies in the précis. This might facilitate an explicit conversation about stakeholder involvement at the beginning of the process to make these strategies implemented more intentionally in practice.

Evaluation Policy Implementation for Reporting Findings

The information gathered through the case studies suggests that many of the reporting decisions lie with the evaluators. In proposals, evaluators are asked to outline what their deliverables will be. In the examples consulted, these were described in broad and vague terms. These examples demonstrate that a priority for the Foundation is the publication of results in academic or peer-reviewed journals. Other types of reporting mentioned in the documents consulted in these case studies include interim and annual reports, policy briefings, presentations, tailored reports to particular communities that receive the intervention

In the complex case, the primary method of reporting has been through peer-reviewed journal articles. The evaluator on the project estimates that they have produced approximately 60 publications related to this project. These publications don't always necessarily relate to evaluating the program. The R&E officer explains, "[The evaluators] publish things that are related to the topic, and use the data that they are collecting. [These articles] don't really evaluate the program. Now [the evaluators] are getting ready to kind of start thinking about what their final evaluation is going to look like, but they still have a few more years." In the meanwhile, despite not being evaluation findings per se, the articles produced are related to the Foundation's idea of building an evidence base around a strategy.

The evaluation team has also submitted interim and annual reports to the Foundation. Beyond the Foundation, and in addition to publishing in peer-reviewed journals, the evaluators have made presentations to the treatment communities, prepared issue briefs that are available on the web, and produced customized summaries for each community based on their specific data.

According to the R&E officer, the evaluation team decides on the venues in which to publish their findings.

In the successful case, according to the evaluator, findings from the implementation study have consisted of, “interim, final report, and three issue briefs, multiple presentations, and a research article for a journal.” The evaluator also added that they report on findings simultaneously to the program and the Foundation part way through the year, and toward the end of the school year. Furthermore, the evaluator explains that “everything we write, we write for the field.” The evaluator also shared his/her ideas about the usefulness of the information produced; s/he says, “I think it was a very rewarding and helpful study for Fun and Games. I can’t tell you how many times [the program] has asked me to present the findings from that work to their own staff nationwide. For years I’ve been traveling around with [the program] to present the findings, so they found them very, very helpful.”

The impact study also shared interim and annual reports with the Foundation. Policy briefs and peer-reviewed articles have also been produced, or are in process. Additionally, several school districts requested specific reports, which were furnished by the evaluators. Finally, much of the reporting has been geared towards policy makers. A Foundation leadership member explains, “we’ll have briefings with policy makers on capitol hill so that they know of the program that it’s maybe in their district if it’s at the national level and what it’s showing with the briefings.” Because of the strength of the design, the Foundation was interested in making a case on a national scale.

The process of reporting findings was also characterized by the collaboration that was a regular feature of this study. The program director shares, “I got very involved when the data were gathered and analyzed and we were making decisions about what to communicate, how and through which channels, so I had a lot to do with that end of it.” More specifically, the evaluators made initial decisions about policy briefs including their contents and structure, and the Foundation then

helped the program take those products and turn them into materials that were more publicly-accessible in terms of length, appearance, and language style.

This process was not without its challenges. The program director further explains, “One of the challenges is that evaluators want to speak very carefully and conservatively about what the findings mean, and we (the program) want to speak very liberally and broadly about what the findings mean. So we had a lot of conflict and discussions over actual words in the briefs where we understood how the audience of school leadership would interpret a phrase, and the evaluators are not as close to that audience.” The Foundation R&E officer played an important role in helping the program and the evaluators determine language that met the needs of both parties.

The overall difficulties experienced in the challenging case extended to the reporting phase. There were internal reports submitted to the Foundation and published on the website of the evaluator, but there were far fewer products submitted for peer review than in the other two evaluations. Each year, the evaluators submitted an annual report to the Foundation, the Evaluation Advisory Panel, and the national program office. The chief product available that summarizes this program is the program results report, which is on the Foundation website. When the final evaluation report was submitted, the evaluators offered to meet with the program office, but that meeting was never held due to exigent circumstances.

Conditions that Support or Inhibit Evaluation Policy Implementation

Throughout this investigation, several stories and examples have emerged that call attention to how the implementation of evaluation policies is impacted by external factors. Of primary importance is how research designs are influenced by federal context and standards surrounding evidence. Each case affords an opportunity to examine different ways in which external factors bear in important ways on the implementation of evaluation policies.

At the inception of this investigation, part of the reason a Foundation made such an appealing setting was its apparent independence from larger state or federal entities that might influence its policies. As the study unfolded, It became clear that despite the fact that the Foundation is not subject to explicit federal oversight, it still exists within a larger context of standards of evidence that ultimately shape its evaluation policies. This plays out primarily in the ways in which evaluation studies are designed.

One of the chief goals of the Foundation's evaluation policies is to influence policy makers. As such, evaluations must produce evidence that will be considered credible to that audience. A research or evaluation must meet certain standards held by the federal government in order to be considered a contribution of evidence to support particular practices. These standards for evidence-based practice manifest in rules that determine what types of studies are published in national repositories that promote interventions that "work." A primary example of this is the What Works Clearinghouse, used to warehouse research that seeks to provide an evidence base to support effective educational interventions. There is a clear preference for randomized control trials (RCT) indicated for inclusion into this database. This is tied into the idea that RCTs are a well-established way to demonstrate causal effects of particular interventions. Even though this idea is the subject of many challenges and constant debate, the fact that it influences federal standards of evidence quality is undeniable.

A preference for a particular type of design is, however, an insufficient condition to support its execution. In only one of the cases presented here, the success case, was an RCT possible. The other two interventions focused on outcomes at a community level. Because communities had to have certain conditions in place in order to be chosen to participate, random assignment was not possible. It worked in the Fun and Games case because schools applied to participate in the

program and the demand exceeded the capacity. These conditions were necessary to support the successful implementation of the RCT.

The challenging case provides an example of a situation in which the desire to influence federal policy proved to be a barrier to successful evaluation. The context surrounding the program suggested to the program director that there was an opportunity to directly influence policy and appropriations through demonstration of an effective program. This program director's desire to capitalize on this moment strongly influenced the design of the evaluation, but ultimately proved to derail the process. The focus on proving that the program worked resulted in a narrow view of the potential for the evaluation. Findings were unable to be used to support this larger political aim, but the myopic view held by the program director obscured the ability of the program and Foundation to learn in other ways from the evaluation.

Finally, in the complex case, an example of when factors within the Foundation, but external to the evaluation, influence the implementation of evaluation policies is shown. This evaluation began following the typical guideline for how evaluations are conducted at the Foundation—the program began with a pilot phase and an implementation study that would influence a demonstration phase and impact study. The officers on the program side of the Foundation, however, were eager to scale the intervention almost immediately. The evaluation had to then depart from the policy and adapt to the changing circumstances of the program. Despite this disruption in policy, it ultimately did not seem to negatively impact the ultimate goals for evaluations. The findings from this study have served to inform both program improvement and the field at large.

CHAPTER 6

DISCUSSION

Introduction

This chapter offers a discussion of the results from the current study, which sought to explore the evaluation policies and their implementation at a philanthropic organization. The following broad research questions guided the investigation:

- 1) What are the evaluation policies of the Robert Wood Johnson Foundation?
- 2) What is the extent to which evaluations are implemented as described by the policy?

The answers to the research questions are positioned in a broader context considering the extant literature on the topic. Furthermore, the implications of the study, the limitations, and directions for future research are discussed.

Review of the Findings

Detailing Organizational Evaluation Policies

The first research question explored the Foundation's evaluation policies and their development and communication. Several important findings are highlighted here. First, the idea of evaluation policies as a system is explored. Then, the ways formalization of evaluation policies might interact with different domains of practice are discussed. Finally, the concepts of "consciousness" and "consequence" (Trochim, 2009) as requirements of evaluation policy are examined.

Evaluation policies as a system. This study first presented a description of the evaluation policies of a Foundation that engages in a significant amount of research and evaluation work. A chief finding is that evaluation policies that indicate the desired outcomes for evaluation practices manifest differently than those that guide evaluation activities. This stands in contrast to how evaluation policies have been theoretically explored in the literature to date. Both the Evaluation

Policy Wheel (Trochim, 2009) and the guidelines offered by the Evaluation Policy Task Force (2007) present a variety of domains that evaluation policies might influence. These are presented categorically; little to nothing is said about the ways in which policies across these domains might interact with one another.

The findings in this study suggest that the relationships between these different domains of evaluation policies are important, and should be discussed in concert with detailing the policies housed therein. This study demonstrates that the goals of the evaluation policies are the basis for the rest of the Foundation's evaluation policies. The operational and goal policies function as a system, rather than distinct units. Accordingly, representing not only the policies, but also the ways in which they are anticipated to interact with one another, is an important step in enhancing thinking about this topic.

A logic model might serve as an instructive example in this case. The goals of the evaluation policies function like desired program outcomes and the operational policies represent the inputs, activities, and outputs that lead to the desired outcomes. The assumptions underlying this model represent the theory of action, or the reasons why the specific activities should lead to those outcomes.

Variable policy formality. Another issue raised in this study is a question of where the tipping point is between formalizing evaluation policies that help further evaluation practice versus over-codifying them such that they restrict the ability of evaluators to be flexible in the field and adaptive to particular contexts. Part of the reason there was little consistency in understanding of the Foundation's operational evaluation policies across records and interviews can be explained by the Foundation's operational premise—it hires and contracts talented experts who will draw on their experience and training to make informed decisions on a case-by-case basis. It does not necessarily serve the Foundation to develop cookie cutter rules for the sake of having a policy that might

ultimately serve to be more restrictive than supportive.

The degree to which policies can and should be formalized might also vary according to the domain of evaluation practice. A greater degree of formalization is appropriate for the more administrative domains like contracting evaluators and reporting findings. These tasks lend better to stricter guidelines because they are not necessarily situations in which policies might restrict the ability to be flexible. Instead, the policies could serve an important communication function in these instances.

It is possible to engage in some speculation about how more formalized policies might contribute to the Foundation's evaluation practice. First, if there were more standardized procedural guidelines, it might help make personnel transitions smoother. In one of the case studies, shifts in personnel introduced challenges because each program officer is allowed to introduce his/her own goals and priorities into a situation. If more clearly written rules guiding the evaluation were in place, it might limit that complication. Second, more formal policies might better help educate and build the capacity of grantees when it comes to evaluation practice. Grantees might better know what to expect during the course of an evaluation, which might ease some of the conflicts that arose over the course of some of the evaluations presented here. Finally, greater formalization of policies would offer more opportunity to discuss and refine them. As Trochim (2009) explains, "Because they are public, written policies, they can be known by everyone in the organization and thus criticized and challenged. Participation and dialogue can occur about which policies make the most sense under which circumstances" (p. 17).

Consciousness and consequence. Finally, Trochim (2009) offers the idea that in order for something to be a policy, it must be adopted with consciousness and consequence. Together, these related concepts describe a situation in which an organization intentionally makes a choice to use policy to guide decisions or actions, and enacts penalties when the policies are not followed. In the

present study, this was a difficult concept to reconcile with the idea that policies can be less formal and more tacitly held (an argument also advanced by Trochim (2009) in the same paper). Very few of the policies elaborated in this study meet these criteria, particularly when it comes to consequence. In no instance was a consequence of not following a policy made apparent over the course of conducting this research.

Even if all the guidelines described in this study fail to meet the criteria to call them “policies,” they still exist and influence evaluation practice in important ways. This calls the idea of this requirement as a necessary condition to warrant attention into question. Despite the lack of consequence for failure to implement policies at the Foundation, some rules and guidelines exist, and bear importantly on the evaluations that are conducted. Findings from this study suggest that even if formal policy does not exist, there might be a middle ground. That is to say that evaluation approaches, when recommended by an organization that funds and supervises evaluation practices, are de facto policies even when consequences for failure to adopt them are not made explicit.

Examining the Implementation of Evaluation Policies

It was difficult in this study to truly capture the strength of evaluation policy implementation. Instead, the findings show examples of how evaluation processes happen in the field. The picture that emerges offers valuable insights into how evaluation policies relate to the organizational functioning of the Foundation.

Organizational learning. The Foundation prides itself on its learning culture. This is evidenced several times in the present study—the Foundation takes great care to learn about its grant-making practices, the effectiveness of the interventions that it funds, and about the field of health and healthcare generally. There are policies within the Foundation that support these learning activities. However, this kind of attention and scrutiny is not directed at its evaluation practice. Perhaps, the Foundation could develop policies that might help support the organizational learning

around the area of evaluation practice.

One of the chief findings about the evaluation policies in this study is that there is a marked difference between policies that are goal statements and policies that are operationalizations of those goals. While there is a tremendous amount of consistency across the Foundation's documents and personnel when it comes to evaluation policy goals, operational decisions of the evaluation policies are decentralized. Each Foundation research and evaluation program officer is able to make evaluation activity decisions based on his/her interests and objectives, and the particular context at hand.

This practice makes sense for the Foundation because it hires and contract experts with extensive training and experience in the evaluation of public health initiatives. Rather than strictly prescribing evaluation activities in a top-down, cookie-cutter fashion, the Foundation trusts that the people involved will call on their extensive training and experience to inform the best design and activities for a given situation. If the Foundation was too prescriptive in its policies, over codification might limit the flexibility and adaptability of evaluations to be responsive to particular conditions.

The upshot, however, is that when learning happens about evaluation practice, it does so in silos. Research and evaluation program officers report that their experiences overseeing evaluations have led to learning about practice. Based on this learning, R&E officers then make changes when overseeing future evaluations. While improving practice based on reflection is important, it is inefficient from an organizational perspective. There does not seem to be a mechanism to support learning in a top-down fashion, or a way to benefit from shared learning.

This is interesting given the structures the Foundation has in place to support its learning in other arenas. For example, when it comes the Foundation strategic objectives, every year the Foundation engages in reporting across several projects. These reports are compiled and prepared

by research and evaluation staff at the Foundation. The process for developing these reports includes reviewing findings from evaluations, and surveying and interviewing key informants. These reports are made public and are discussed internally at the Foundation. The reports are designed to help the Foundation learn in a broader way about its progress towards reaching its strategic goals.

When it comes to research and evaluation on the program level, the same research and evaluation team, as a part of completing the program results reports, also issues grant making lessons, and when warranted, lessons learned about evaluation. The main difference here is that these lessons are not circulated with intention beyond those directly involved with the project. Much of the experiential knowledge gained from participating in research and evaluation lies with one individual in the research and evaluation department. The person chiefly responsible for compiling these reports indicated that s/he frequently attempts to offer advice based on what s/he has discovered to others when are engaging in similar work. This happens informally though, not as a matter of course.

Perhaps evaluation policies could serve the function of institutionalizing this learning. For this to happen, structures would need to be put in place. These structures would include things like having an internal meeting at the conclusion of a project during which lessons are broadly discussed. These lessons can then be synthesized, and shared Foundation wide. The support of organizational learning is an area in which having more explicit evaluation policies might help. Trochim (2009) explains that only when policies are made explicit can they be debated and ultimately improved upon. Evaluation policies might serve as the objects of a feedback loop in which the efficiency of learning about evaluation practice within the Foundation could be improved.

Diversity and innovation. Another theme that emerged during conversations primarily with Foundation leadership is the idea of diversity. Concern was expressed that there is not enough diversity in terms of ethnicity, socioeconomic status, etc. within the pool of evaluators. This

particularly troubled one individual, partly because s/he feels that it is important for evaluators to match the population they are evaluating. Beyond characteristic diversity of evaluators, there is also an issue with limiting the diversity of ideas. The current Foundation policies may, in fact, serve to limit, rather than promote innovative thinking when it comes to evaluation practice.

To see how this works, an example is found in the processes by which evaluators are selected and contracted. As a reminder, there are two methods by which evaluators are selected for a project. The first is through sole sourcing. In this process, the research and evaluation officer selects an individual or firm with whose work and content expertise they are already familiar. When this happens, the research and evaluation program officer frequently consults with their colleagues and relies on a pool of knowledge rather than just their own experience. This method of contracting evaluators is partly motivated by the desire (and an implicit policy) that evaluations should start as programs are beginning. The fastest way to do this is to contact with an evaluator directly.

The second method of contracting evaluators is through an invited bid. In this situation the R&E program officer develops a request for proposal (RFP) and invites preselected evaluators or firms to reply. The list of evaluators is often determined in the same way as the sole sourcing procedure—R&E officers are reliant on their own experiences, or consultations with Foundation colleagues to determine who will be asked to reply to the RFP. An interesting note here is that Foundation is currently working on building evaluator database to help facilitate this process. However, this is not yet widely used, and still involves significant prescreening on the part of the Foundation.

The result is that the Foundation is dealing with a limited pool of evaluators. Again, this practice makes sense. As already mentioned, one program officer explains, “this isn’t amateur hour.” Because the Foundation is operating on such a major scale and with significant resources, it is unwilling to risk having someone unknown engage in such important work. There is a trade-off,

however. Continually going back to the same people means that there might be a lack of diversity in ideas.

This is exacerbated by the fact that research design ideas almost always originate with the Foundation R&E officer. Regardless of whether or not the evaluators are sole-source or selected through competition, the R&E officer initiates the process by drafting research and evaluation questions. It was very consistently reported that there is still a lot of back-and-forth involving the opinions of the evaluator and the program in finalizing the questions, but there is no evidence the questions ever changed dramatically. The situation that results is one in which a similar group of people are asking a similar set of questions, just across different contexts.

The lack of diversity in evaluation questions and approaches, however, is not terribly problematic. The Foundation does exceptional work in the field of public health that is in many ways supported by the rigorous evaluations it commissions. However, there is an unrealized potential here in which the Foundation could be doing more to further the field of evaluation and advance practice. Given what was said at the beginning of the study—that Foundations play an important role in allowing the ability to test new things—we should be able to extend this analogy to the practice of evaluation. The Robert Wood Johnson Foundation is a perfect setting to explore innovative ideas and approaches.

Again, the fact that the Foundation is operating in such a high-stakes arena does demand certain amount of prudence. Perhaps a solution might be found in structuring evaluation policies that differ according to the size of the intervention at hand. The Foundation could maintain its current practice for large-scale interventions through which it is trying to affect public policy. A different set of rules might then exist for evaluations of programs under certain amount of money. At present, programs under \$400,000 do not always have evaluations associated with them, and require less organizational approval. Perhaps a policy could be written that in those cases,

evaluations are contracted through an open bid process in which there are few, if any, preset guidelines. This would offer evaluators new to the Foundation the opportunity to work with it, and allow for innovative evaluation approaches to be tested in real-world settings (Trochim, 2009). Eventually, if the Foundation is satisfied with the work of these evaluators, they could be added to the evaluator database. Over time, the diversity of that database would increase as the pool of potential that evaluators expands.

Implications

This study has several important implications for the field. Specifically, implications for evaluation practice and agencies that commission and oversee evaluation work are discussed below.

Evaluation Practice

This study has several implications for evaluation practice, for both practitioners of evaluation and for the field in general.

Even though individuals contracted to conduct evaluations will have limited, if any, ability to impact the policies that guide their work, having an awareness of the ways in which evaluation policies function and influence their practice might help enhance it. Specifically, an understanding of the evaluation policy goals held by an organization may help evaluators better respond to RFPs. Furthermore, understanding how prescribed activities are supposed to lead to goals would provide a framework for strengthening proposals, or perhaps an opportunity to challenge particular assumptions or decisions that seem to be policy directives.

Secondly, evaluation policies could serve as a mechanism to promote innovation in evaluation. The established standards of practice are prevalent in the evaluation policies witnessed here. In order to evolve as a field, opportunities to test new methods and approaches in real-world settings are necessary. However, challenging the status quo is difficult; perhaps even more so in evaluation than research. Greater amounts of flexibility and researcher-initiated questions and

processes frequently characterize the latter. Evaluation is frequently more restricted because of its immediate connection to policy makers and practitioners. The stakes for use of findings are often higher, and as a result, there may be less willingness to engage in risk-taking by organizations that are funding the evaluation efforts. It seems as if innovation would be more prevalent if there were policies specifically designed to promote it. Large foundations, like Robert Wood Johnson, could do a great deal in developing conditions through evaluation policies that support exploration and testing of new evaluation ideas and approaches.

Agencies that oversee evaluations

Beyond philanthropic organizations, there are many agencies that are responsible for commissioning and overseeing evaluation work. Pressure to prove that interventions are effective has increased in the current era of accountability. While the Robert Wood Johnson Foundation was an excellent site to conduct this investigation because of its strong commitment to research and evaluation and its recognized expertise in that arena, it proves the exception rather than the rule. Frequently, other agencies (public or private) are thrust into an expectation to oversee evaluation without any expertise to do so. Evaluation policies, developed through expert consultation, might assist these organizations in conducting efficient and valuable evaluation work.

Limitations

As with any study, this investigation has limitations worthy of mention. The methodological, analytical, and logistical challenges met in this study are discussed in turn.

Methodological

Qualitative inquiry, by nature, has several limitations. First, it is dependent on the researcher as the instrument of data collection and analysis, and as such, is subject to the researcher's biases and particular ways of viewing the world. Several steps were taken to address this. First, different types of data sources within each phase of the study were consulted in an effort to triangulate findings.

Second, the analysis is influenced by previous scholarship in an effort to help link the conclusions drawn here to extant theories.

Of primary concern is the ability to generalize these findings both within the Foundation's practices and beyond. The selection of three cases, while strategic, represents a very limited sample of the Foundation's evaluations over its more than 40-year history. The cases present a snapshot of how policies were enacted at a particular moment within a particular context, and it is challenging to generalize these findings to the practices of the entire Foundation. This difficulty was anticipated at the beginning of the research and partly served to inform the selection of the cases. While not representative, the diversity of circumstances across the three cases affords an examination of how policy implementation is influenced by a variety of settings.

The challenge of generalizability persists in considering the ways in which this study might speak to considerations of evaluation policies at other organizations. While The Robert Wood Johnson Foundation's significant resources, long history with inquiry, and commitment to research and evaluation make it an ideal setting for this study, it is a unique organization. At the granular level, it is difficult to transfer these findings beyond the Foundation. However, the larger conclusions shed insight into how evaluation policies, broadly, might help influence evaluation practice.

Analytical

Analysis presents several challenges to this investigation. As with data collection, in the analysis phase, the researcher is the primary instrument and as such, the interpretations are influenced by the researcher's biases and experiences. This was largely addressed by consulting multiple data sources and previous scholarship relevant to the topic.

The Foundation does not have clearly articulated evaluation policies—they were distilled in this study through consultation of a large body of evidence. As an outsider to the Foundation, the

researcher's ability to draw inferences that accurately represent Foundation policies is limited. This was a challenging activity filled with both consistencies and contradictions. To address this, both confirming and disconfirming evidence are presented throughout the findings chapters. Additionally, the Foundation staff was regularly consulted as questions arose.

Logistical

Finally, this study presents several logistical challenges. Specifically, the Foundation served as a primary gatekeeper for the data, and some of the experiences that interview participants were asked to reflect upon happened many years ago.

The Foundation was very supportive of this work, and assisted a great deal in selecting data for analysis and scheduling all interviews. This study relied on Foundation staff for access to its internal documents and for access to interview participants. This was essential to complete the investigation, but it does leave the study open to questions about whether this data might have been in some way specifically selected by the Foundation so that its practices could be reflected in the best possible light. However, this concern is unwarranted, partly because of the selection of the challenging case. The fact that this case was not a reflection of the Foundation's best efforts was widely acknowledged by both people who worked at the Foundation during that time, and involved parties external to the Foundation.

Furthermore, the Foundation is very proud of its organizational culture as it relates to learning. Its willingness to host this work provides further evidence to support the idea that RWJF is interested in gathering information, regardless of whether or not the findings are positive.

Preliminary findings from this study were presented to the Foundation in March of 2014. RWJF leadership members were very receptive to criticisms of the Foundation's practice, and engaged enthusiastically with the findings.

The final logistical challenge in this investigation has to do with the age of one of the case

studies. The project concluded several years ago, so in interviews people were asked to recall old information. In several cases, participants did not remember the specifics. Fortunately, interview responses were corroborated by document analysis. Ultimately the case does not suffer from inaccuracies due to recall challenges. Finally, there was one instance in which a potential participant did not respond to requests for interviews. S/he was the program director for the challenging case, and his/her perspective would have been valuable to have. Despite this challenge, the program director's perspective is represented through some of the reporting that s/he submitted to the Foundation.

Directions for Future Research

In addition to shedding light on how evaluation policies might work to shape evaluation practice, the findings from this study suggest both theoretical and applied areas in which further investigations might expand understanding. Some of these are outlined below.

Theoretical Investigations

Our understanding of evaluation policies could be enhanced through more theoretical investigations. Thus far, much of the work in the field about evaluation policies has focused on the various aspects of practice that they might govern. Findings from this study suggest that the field might benefit from more probing into how the specific areas of evaluation practice might interact with more explicit evaluation policies. The issue overarching this inquiry concerns striking a balance between codifying policies, yet allowing for enough flexibility to support adaptations in the field when they are warranted.

As previously discussed, one of the key issues surrounding successful implementation policies is that they strike a balance between providing guidelines and allowing for enough flexibility in the field. Future research in this area might help improve understanding of the continuum between over codifying policies and a complete lack of rules and guidelines. One potential way this

could be investigated further is through a Delphi study, a process by which expert opinion is solicited in an iterative process to reach consensus about a topic (Helmer, 1967). The Delphi study is conducted via survey anonymously. A Delphi study could afford an opportunity to attain expert consensus among evaluation practitioners and scholars to suggest the degrees to which evaluation policies should be made explicit. More recently, Delphi studies have been adapted for face-to-face techniques. Both methods might be appropriate for this type of investigation.

Relatedly, these findings suggests that the degree of codification might vary according to which aspect of evaluation practice the policy intends to govern. For example, RWJF practices regarding the selection and contracting of evaluators might benefit from more stringent rules and guidelines. However, certain aspects of evaluation practice, like data collection and stakeholder engagement, are sometimes too context-dependent to be able to adhere to overly explicit rules. The field could benefit from a simulation study in which evaluators could be assigned to two separate conditions—one where the evaluation is guided by very explicit policies over certain areas, and the other in which fewer rules are imposed. The evaluators could then be asked to design evaluations. Results from the two groups could be compared to see in what ways policies impact evaluation designs.

Applied Investigations

The work presented in this study also points to more applied areas of future research. One option is to conduct a similar investigation into the evaluation policies of other types of organizations that commission evaluations. Conducting a similar study in a different philanthropic or non-profit organization might help to replicate these results. Furthermore, attention could be turned to federal or state bodies that commission evaluations to build understanding of what evaluation policies mean for practical work.

Ultimately, an investigation that more closely examines the implications of evaluation

policies is needed. Beyond questions of how the existence or absence of evaluation policies might impact practice, further study is necessary to uncover the ways in which specific types of policies interact with evaluation activities. This is of critical importance because of the hypothesized connection between policies, practice, and the data used to make high-stakes decisions about programs.

Beyond explorations of how existing policies influence practice, attention could also be turned to the potential power of policies to shape evaluation practice as a field. In this study, speculation is offered about how policies might be able to promote innovation in practice. The field could benefit from further exploration into this aspect of evaluation policies.

Final Remarks

This study was designed to shed light on a Foundation's evaluation policies in an effort to draw connections about how evaluation policies influence evaluation practice. The investigation helped further understanding of these connections, and generated some insight into the components of evaluation policies, and the role they might play in shaping the future of evaluation practice.

Appendix A
Recruitment Letter

DATE

NAME,

The Robert Wood Johnson Foundation wants to shed light on how its policies and procedures affect the evaluations that it commissions. A major interest is the alignment between evaluation policies and how they are understood and implemented by evaluators and program personnel. The Foundation will use the study results to inform the development of more explicit guidelines for its evaluation work.

Lisa Dillman of UCLA is leading this effort. As part of this work she is undertaking several case studies of evaluations funded by the Foundation.

We are contacting you so that Ms. Dillman can schedule a brief telephone interview to discuss your work with the Foundation and the *NAME OF PROGRAM* program.

A consent form describing the study in further detail is attached. If you have questions regarding this study, please feel free to contact Lisa Dillman (ldillman@ucla.edu) or Denise Herrera, Evaluation Associate at the Robert Wood Johnson Foundation (dherrera@rwjf.org).

Thank you for your time and consideration.

Sincerely,



Laura C. Leviton, Ph.D.
Senior Advisor for Evaluation

Cc Lisa Dillman, Denise Herrera

Appendix B

Interview Consent form

University of California, Los Angeles

CONSENT TO PARTICIPATE IN RESEARCH

Alignment between Intention and Implementation: A Case Study of Evaluation Policy

Lisa Dillman, under the advisement of Dr. Christina Christie from the Graduate School of Education and Information Studies at the University of California, Los Angeles (UCLA) is conducting a research study.

You were selected as a possible participant in this study because you work in or with the Robert Wood Johnson Foundation. Your participation in this research study is voluntary.

Why is this study being done?

This study is designed to shed light on the evaluation policies of the Robert Wood Johnson Foundation. The purpose of this study is to document both the intended and enacted evaluation policies and procedures.

What will happen if I take part in this research study?

If you volunteer to participate in this study, the researcher will ask you to do the following:

- Participate in an interview with the researcher that will cover topics about your roles and responsibilities related to the evaluation practices of the Foundation. Additionally, you will be asked to describe your understanding of the Foundation's evaluation policies.

How long will I be in the research study?

Participation will take a total of about 30 minutes to one hour for the initial interview. If follow-up is required, it will take place several months after the initial interview over the phone. Follow-up will last approximately 30 minutes.

Are there any potential risks or discomforts that I can expect from this study?

There are minimal risks associated with participation in this study. You may experience discomfort in answering some of the interview questions.

Are there any potential benefits if I participate?

There are no anticipated individual benefits from participating in this study. The results of the research will be used to inform discussions about the importance of evaluation policy as it relates to the practice of evaluation.

Will information about me and my participation be kept confidential?

Any information that is obtained in connection with this study and that can identify you will remain confidential. It will be disclosed only with your permission or as required by law. Each participant will be assigned a code that the researcher will use to identify his/her interview transcript in cases where a follow-up interview is deemed appropriate. Interview transcripts will be kept on the researchers password-protected personal computer. The researcher will be the only person with access to the codes. Once the data analysis is completed, the codes will be destroyed.

Data obtained through the interviews will be reported mostly in the aggregate. In cases where direct quotes from individuals are used, the researcher will remove all information that could make the quote identifiable.

What are my rights if I take part in this study?

- You can choose whether or not you want to be in this study, and you may withdraw your consent and discontinue participation at any time.
- Whatever decision you make, there will be no penalty to you, and no loss of benefits to which you were otherwise entitled.
- You may refuse to answer any questions that you do not want to answer and still remain in the study.

Who can I contact if I have questions about this study?

- **The research team:**

If you have any questions, comments or concerns about the research, you can talk to the one of the researchers. Please contact:

Lisa Dillman
310-994-9465
ldillman@ucla.edu

Christina (Tina) Christie
310-825-0432
tina.christie@ucla.edu

- **UCLA Office of the Human Research Protection Program (OHRPP):**

If you have questions about your rights while taking part in this study, or you have concerns or suggestions and you want to talk to someone other than the researchers about the study, please call the OHRPP at (310) 825-7122 or write to:

UCLA Office of the Human Research Protection Program
11000 Kinross Avenue, Suite 211, Box 951694
Los Angeles, CA 90095-1694

You will be given a copy of this information to keep for your records.

SIGNATURE OF STUDY PARTICIPANT

Name of Participant

Signature of Participant

Date

SIGNATURE OF PERSON OBTAINING CONSENT

Name of Person Obtaining Consent

Contact Number

Signature of Person Obtaining Consent

Date

Appendix C

Interview Protocol: Foundation Leadership

Questions:

- As I briefly described in my email, I am doing my dissertation on evaluation policy. I am using a very broad definition – basically any rule or principal that a group or organization uses to guide its decisions and actions when doing evaluation. Given that definition, what is your vision of evaluation policy at the Foundation?
- Given your role in the Foundation, I think you may be able to help me understand the idea of evaluation policy in the context of policies that are Foundation-wide. Considering the policy question Foundation-wide, how are policies developed and determined? Who is involved? What is your role in developing and shaping policy?
- Something that has come up quite a bit in these conversations is the idea that there is a culture here that is incredibly supportive of research and evaluation. People have pointed to the Foundation’s origins with Johnson and Johnson as a key reason for this. It got me thinking about the difference between culture and policy (culture being what everyone does and no one talks about and policy being explicit rules or guidelines). Can you talk about how you see the two things playing out at the Foundation?
- I know you led the health group and were a senior program officer prior to moving in to your current role. Will you tell me a little about how you have seen the research and evaluation work change over time?
- Can you think of examples of particularly outstanding research or evaluation activities? What makes them that way? Is it about the report itself or do external factors influence it?
- How about the other side, when things don’t go particularly well. What features do they have in common? What is the Foundation’s response when things don’t work?
- In a general way, how do you know when an evaluation is “good” or “bad?” What are the key factors that go into distinguishing that?
 - Particular methods
 - The way conclusions are reached
 - The way results are communicated/received
 - How might these ideas relate to the Foundation’s policies around evaluation?
- In cases where something stands out, either for being particularly helpful, or particularly problematic, what is the response? What do you think the ideal response should be?
 - Is it a determination you make at the end of the day when the report is filed, or can you see problems or glimmers of brilliance as the process unfolds?
 - To what extent do you think these instances are a reflection of evaluation policies?

Appendix D

Interview Protocol: Program Leadership

Questions:

- To get started, I was hoping you could tell me a little about your day-to-day work with the Health Group, and specifically how it relates to the work done in the research and evaluation division?
 - Follow ups might include evaluation coordination, use, etc.
- According to your bio you see your role as “contributing to the Foundation’s intellectual and organizational development, and managing program operations to ensure that we meet RWJF’s goals of reversing the childhood obesity epidemic, driving fundamental improvements in the nation’s public health system, and addressing the needs of the country’s most vulnerable populations.” How does research and evaluation fit into that?
- I’m asking everyone generally to describe the evaluation policies of the Foundation. I am talking about evaluation policy in a very broad sense – any rule or principal that a group or organization uses to guide its decisions and actions when doing evaluation. Will you tell me what comes to mind about evaluation policies at Robert Wood Johnson?
 - Probe for specifics – beyond goals, what about particular methodologies,
- I was told last week of a somewhat recent shift in the timeline for evaluations in that now they are designed alongside programs. Given your long tenure with the Foundation, I assume you have seen both sides of this process. Can you talk a little about what this change has meant for you and your work with the Foundation?
- How do you make decisions about what gets evaluated and when?
 - How do you coordinate with the Research and Evaluation Program Officers about that?
 - Goals
 - Methods
 - Ideas about designs
 - CFPs
 - Once an evaluation of a program is under way, what is your involvement with it?
 - How do you use evaluation findings when all is said and done to support your work?
- Aside from being a lawyer by training, you work a lot on the legal policy side of things. Will you talk a little about how research and evaluation fits in to that area of the work of the Foundation?
- I’m sure there is a range of quality evidenced in the evaluations you have seen. What do you think the key factors are that go into determining whether an evaluation is good or bad? Or

you can think about it in terms of useful or helpful for you. In what ways do you think that might relate to evaluation policy?

Appendix E

Interview Protocol: R&E Program Officer/Senior Program Officer

Questions:

- To get started, I was hoping you could tell me a little about your day-to-day work with the Foundation, and specifically how it relates to the work done in the research and evaluation division?
Follow ups might include evaluation coordination, use, etc.
- Based on response:
Do you evaluate responses to CFPs?
What are some key factors you look for in an evaluation design?
- How do you determine what will make a good evaluation?
How do you use evaluation findings?
- I'm asking everyone generally to describe the evaluation policies of the Foundation. I am talking about evaluation policy in a very broad sense – any rule or principal that guides a group or organization uses to guide its decisions and actions when doing evaluation. Will you tell me what comes to mind about evaluation policies at Robert Wood Johnson?
 - Probe for specifics – beyond goals, what about particular methodologies,
- What is your role in relation to RFPs/CFPs? If any – what do you think they communicate about evaluation policies?
- I'm sure there is a range of quality evidenced in the evaluations you have seen. What do you think the key factors are that go into determining whether an evaluation is good or bad? In what ways do you think that might relate to evaluation policy?
- What aspects of evaluation do you think would benefit from having policies? (design, stakeholder involvement, methods, reporting, etc.)
 - Does that relate to eventual use of evaluation findings? In what ways?
- How do you (not necessarily you, but the R&E department) make sure that policies are followed?
- If you had to choose several standout evaluations and several that failed to meet expectations, which would they be? On what criteria would you base those decisions? How much of their success or failures do you think relates to evaluation policies?

Appendix F

Case Study: Program Personnel Interview Protocol

Program Background

- When did your involvement with SPECIFIC PROGRAM start?
 - What was your role?
- What are the goals of the program?
- Were you involved in developing the evaluation contract or selecting the evaluator?
 - Did you know the evaluators previously?
 - How was the evaluator selected/contracted?

Evaluation Design/Questions

- What are the goals for the evaluation?
- What kind of role, if any, did you play in the evaluation design?
- What did you hope to learn from the evaluation?

Communication/Foundation Relationship

- Will you tell me a little about your interactions with the evaluator and/or the RWJF Research and Evaluation Program Officer?
- How would you describe your relationship with the evaluators?
- How would you describe your relationship with the RWJF Research and Evaluation Program Officer?
- Will you tell me about a time a challenge was encountered with the evaluation?
 - How do you go about resolving these issues when they come up?

Stakeholders

- Who do you consider to be the key stakeholders in the evaluation?
- To the best of your knowledge, in what ways were they engaged in this study?

Findings/Reporting

- You talked earlier about the goals for the program. Please describe how the evaluation has helped you understand the ways in which the program met its goals.
 - Describe some of the challenges encountered along the way.
- How have evaluation findings been reported?

Use

- Who would you identify as the primary users of the evaluation?
- In what ways have the findings for this evaluation been used? By whom and for what purposes?
- In what ways has the evaluation contributed to meaningful changes in the program?
- In what ways has the evaluation contributed to meaningful changes in Foundation practices?

Appendix G

Case Study: Program Officer Interview Protocol

Program Background

- When did your involvement with SPECIFIC PROGRAM start?
 - What was your role?
- What are the goals of the program?
- Were you involved in developing the evaluation contract or selecting the evaluator?
 - Did you know the evaluators previously?
 - How was the evaluator selected/contracted?

Evaluation Design/Questions

- What are the goals for the evaluation?
- What kind of role, if any, did you play in the evaluation design?
- What did you hope to learn from the evaluation?

Communication/Foundation Relationship

- Will you tell me a little about your interactions with the evaluator and/or the RWJF Research and Evaluation Program Officer?
- How would you describe your relationship with the evaluators?
- How would you describe your relationship with the RWJF Research and Evaluation Program Officer?
- Will you tell me about a time a challenge was encountered with the evaluation?
 - How do you go about resolving these issues when they come up?

Stakeholders

- Who do you consider to be the key stakeholders in the evaluation?
- To the best of your knowledge, in what ways were they engaged in this study?

Findings/Reporting

- You talked earlier about the goals for the program. Please describe how the evaluation has helped you understand the ways in which the program met its goals.
 - Describe some of the challenges encountered along the way.
- How have evaluation findings been reported?

Use

- Who would you identify as the primary users of the evaluation?
- In what ways have the findings for this evaluation been used? By whom and for what purposes?
- In what ways has the evaluation contributed to meaningful changes in the program?
- In what ways has the evaluation contributed to meaningful changes in Foundation practices?

Appendix H

Case Study: Research & Evaluation Officer Interview Protocol

Program Background

- When did your involvement with SPECIFIC PROGRAM start?
 - What was your role?
- Will you describe your involvement in the evaluation contract?
 - Did you know the evaluators previously?
 - How was the evaluator selected/contracted?

Evaluation Design/Questions

- What are the goals for the evaluation?
- Please describe briefly the evaluation design?
- Who has been involved in designing the evaluation?
- To what extent was your design a “negotiation” between you and the evaluator?
 - How was the design finalized –describe the process
- How did the design change overtime?
 - Who and why initiated these changes?

Communication/Foundation Relationship

- Will you tell me a little about your interactions with the evaluator thus far?
- How would you describe your relationship with the evaluators?
- Will you tell me about a time a challenge was encountered, or when something about the evaluation didn’t go as anticipated?
 - How do you go about resolving these issues when they come up?

Stakeholders

- Who do you consider to be the key stakeholders in the evaluation?
- To the best of your knowledge, in what ways were they engaged in this study?

Findings/Reporting

- You talked earlier about the goals for the evaluation. Please describe the ways in which the evaluation has met these goals.
 - Describe some of the challenges encountered in meeting the goals
- How are findings reported?
 - Aside from the Foundation, to whom else are findings reported? (public, peer review, program, other stakeholders)

Use

- Who would you identify as the primary users of the evaluation?
- In what ways have the findings for this evaluation been used? By whom and for what purposes?
- In what ways has the evaluation contributed to meaningful changes in the program?
- In what ways has the evaluation contributed to meaningful changes in Foundation practices?

Appendix I

Case Study: Evaluator Interview Protocol

Selection

- When did your involvement with NAME OF PROGRAM start?
- How were you selected as the evaluator for NAME OF PROGRAM?
 - Had you worked with the Foundation, or any of the program officers previously?
 - If so, how would you describe your relationship?

Evaluation Design/Questions

- What were the goals for the evaluation?
- Please describe briefly the evaluation design.
- What questions did your design intend to address?
 - What did you hope to learn?
 - To what extent were your expectations aligned with the Foundation's?
- What informs your design/method choices?
- Who was involved in designing the evaluation?
- To what extent was your design a “negotiation “ between you and the Foundation?
 - How was the design finalized – please describe the process.

Communication/Foundation Relationship

- Once the design was finalized, what role did the Foundation play in your evaluation?
- Will you tell me a little about your interactions with the Foundation?
 - What was the ongoing communication like with the Foundation?
- How would you describe your relationship with program officers and R&E officers?
- Will you tell me about a time when you encountered a challenge, or something about the evaluation didn't go as anticipated?
 - How did you go about resolving these issues when they come up?

Stakeholders

- Who do you consider to be the key stakeholders in this study?

- In what ways have stakeholders been engaged in this study?

Findings/Reporting

- Please describe the ways in which the evaluation has met its goals.
 - Describe some of the challenges encountered in meeting the goals.
- Describe your approach to reporting.
 - To whom do you report findings?
 - How do you report to these groups?

Use

- Who would you identify as the primary users of the evaluation?
- In what ways were the findings for this evaluation used? By whom and for what purposes?
- In what ways did the evaluation contributed to meaningful changes in the program?

Appendix J

Evaluation Policy Coding Framework

CODE	DEFINITION
Evaluation Policy Contexts/Inputs	The circumstances within which evaluation occurs.
Program	What does the policy indicate about the program in advance of conducting an evaluation?
Evaluation	
Identification	Which programs or entities are the objects of evaluation?
Definition	How is evaluation defined?
Schedule	How often are evaluations scheduled? When should they be scheduled? How are the various components scheduled?
Evaluator	
Selection	How are evaluators selected?
Technical Expertise	What technical or methodological expertise/training should the evaluator have?
Content Knowledge	What substantive knowledge should the evaluator have?
Interpersonal Skills	What interpersonal skills should the evaluator poses?
Values	What would be the values, beliefs, priorities, orientation, and attitudes of the evaluator?
Other	What other characteristics of the evaluator are needed?
Stakeholders	
Identity	What stakeholders should be involved in the evaluation? How should these stakeholders be identified or chosen?
Inclusion Reason	What is said about why stakeholders should be included in the evaluation? What are the reasons one might want to involve stakeholders? How is their inclusion justified?
Resources	
Personnel	Who are the personnel involved in the evaluation?
Time	What is said about how much time is/should be devoted to the evaluation?
Financial	What is said about finances/budgeting for evaluation?
General	What is said, broadly, about resources devoted to evaluations?
Evaluation Activities	The procedures used in planning and carrying out evaluations.
Capacity building	What is said about efforts to build the capacity of program personnel to participate in/conduct evaluations?
Planning/Preparation	
Contracts	What is said about the formal contracts between evaluators, the Foundation and program personnel?
Understanding the Program	What sorts of activities are implemented in order to help the evaluator and stakeholders understand the program?
Designing the Evaluation	What sorts of activities are involved in designing the evaluation (including decisions about study questions, methods, etc.)?

Roles	How are roles of involved parties differentiated?
Conducting the Evaluation	
Methods	What does the policy communicate about methods to be used in evaluation?
Evaluator's Role	What is the role of the evaluator in conducting the evaluation? What are her/his responsibilities?
Foundation's Role	What activities does the Foundation undertake during the evaluation?
Program's Role	In what ways is the program to be involved in the evaluation?
Administration	What is said about the logistics of coordinating and administering the evaluation?
Communication	What is said about communication between parties involved in the evaluation?
Managing Relationships	What does the policy indicate about how relationships between parties should be managed?
Stakeholder Participation	In what ways (and to what extent) should stakeholders be involved in the evaluation? What are their roles and responsibilities?
Data Collection	What is the approach to data collection? Does this approach emphasize particular sources of data or collection methods? Who is involved in collection?
Data Analysis	What's the approach to data analysis? Are there particular methods that are typical of the approach? Who is involved data analysis?
Developing Conclusions	How are conclusions reached? What is the basis for a conclusion, and whose is it?
Findings	What is said about evaluation findings/results?
Ethics	What is noted about addressing ethical issues in evaluation?
Reporting	
Format	Through what means are evaluation findings communicated?
Contents	What information is included in evaluation findings?
Audience	Who is the intended audience for evaluation findings?
Metaevaluation	What are the practices described for evaluating evaluations?
Evaluation Effects/Consequences	The desired outcomes of the evaluation policy.
Goals	What does the evaluation hope to accomplish?
Findings	What is said about evaluation findings? Note: this is just about findings generally, not about their use. What specifically results from the evaluation?
Use	How are evaluation findings used?
External Factors	Are there factors or conditions (external to the evaluation itself, beyond its influence) that could affect the extent to which the evaluation achieves the intended effects? May be thought of as confounding variables.
Assumptions	The underlying beliefs that inform/motivate the proposed

	evaluation policy.
Program	What assumptions are made concerning programs that are the object of evaluations?
Epistemology	What assumptions are made concerning knowledge or reality?
Stakeholders	What assumptions are made concerning stakeholders
Evaluation	What assumptions are made about evaluation generally?
Methods	What assumptions are made about the ways of gathering information over the course of an evaluation?
Results	What assumptions are made about evaluation results?
Reporting	What assumptions are made about reporting or otherwise communicating findings?
Use	What assumptions are made about the use of findings?

References

- Alkin, M. C. (2011). *Evaluation Essentials from A to Z*. New York, NY: Guilford Press.
- American Evaluation Association. (2007). AEA Evaluation Policy Task Force Charge. Retrieved from <http://www.eval.org/EPTF.charge.asp>
- Behrens, T. R., & Kelly, T. (2008). Paying the piper: Foundation evaluation capacity calls the tune. *New Directions for Evaluation*, 2008(119), 37–50. doi:10.1002/ev.267
- Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation*, 1987(33), 5–18. doi:10.1002/ev.1443
- Carman, J. G., Fredericks, K. A., & Introcaso, D. (2008). Government and accountability: Paving the way for nonprofits and evaluation. *New Directions for Evaluation*, 2008(119), 5–12. doi:10.1002/ev.265
- Chen, H.-T. (2004). The roots of theory-driven evaluation: Current views and origins. In M. C. Alkin (Ed.), *Evaluation Roots: Tracing Theorists' Views and Influences* (pp. 132–152). Thousand Oaks, CA: Sage Publications, Inc.
- Datta, L. (2009). Golden is the sand: Memory and hope in evaluation policy and evaluation practice. *New Directions for Evaluation*, 2009(123), 33–50.
- Flynn, P., & Hodgkinson, V. A. (2001). Measuring the Contributions of the Nonprofit Sector. In P. Flynn & V. A. Hodgkinson (Eds.), *Measuring the Impact of the Nonprofit Sector* (pp. 3–16). New York, NY: Kluwer Academic/Plenum Publishers.
- Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry*, 36(6), 717–732.
- Hall, P. D. (2004). A Historical Perspective on Evaluation in Foundations. In M. T. Braverman, N. A. Constantine, & J. K. Slater (Eds.), *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy* (pp. 27–50). San Francisco, CA: Josey-Bass.

- Hansen, M., Alkin, M. C., & Wallace, T. L. (2013). Depicting the Logic of Three Evaluation Theories. *Evaluation and Program Planning*. doi:10.1016/j.evalprogplan.2012.03.012
- Helmer, O. (1967). *Systematic Use of Expert Opinions*. RAND Corporaion.
- Mark, M. M., Cooksy, L. J., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. *New Directions for Evaluation*, 2009(123), 3–11.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis*. Thousand Oaks, CA: Sage Publications, Inc.
- National Center for Charitable Statistics. (2012). *Quick Facts About Nonprofits*. Retrieved February 4, 2013, from <http://nccs.urban.org/statistics/quickfacts.cfm>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2001). Retrieved from <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, 11(4), 327–344.
- Oberfield, A., & Dees, J. G. (1992). *Note on Starting a Nonprofit Venture*. Boston, MA.
- Ostrower, F. (2004). *Attitudes and Practices Concerning Effective Philanthropy Survey Report*. Washington, DC: The Urban Institute Press.
- Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. Thousand Oaks, CA: Sage Publications, Inc.
- Porter, M. E., & Kramer, M. R. (1999). Philanthropy's new agenda: Creating value. *Harvard Business Review*, 121-130.
- Robert Wood Johnson Foundation. (2010). *Research and Evaluation*. Princeton, NJ. Retrieved from <http://www.rwjf.org/content/dam/files/rwjf-web-files/Framing-Strategy/RE%20Framing%20Doc%20061110.pdf>

- Robert Wood Johnson Foundation. (2012). *Celebrating 40 Years of Research and Evaluation*. Princeton, NJ. Retrieved from http://www.youtube.com/watch?feature=player_embedded&v=l-WC32_8Uis
- Roeger, K. L., Blackwood, A. S., & Pettijohn, S. L. (2012). *The Nonprofit Almanac 2012*. Washington, DC: The Urban Institute Press. Retrieved from <http://www.urban.org/books/nonprofit-almanac-2012/>
- Salamon, L. M. (2001). What is the Nonprofit Sector and Why Do We Have It? In S. J. Ott & L. A. Dicke (Eds.), *The Nature of The Nonprofit Sector* (pp. 162–166). Boulder, CO: Westview Press.
- Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage Publications, Inc.
- Stake, R. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage Publications.
- Stevhan, L., King, J., Ghore, G., & Minnema, J. (2005). Establishing Essential Competencies for Program Evaluators. *American Journal of Evaluation*, 26(1), 43–59.
- Taylor-Powell, E., & Henert, E. (2008). Developing a logic model: Teaching and training guide. the Board of Regents of the University of Wisconsin. Retrieved from <http://www.uwex.edu/ces/pdande/evaluation/pdf/lmguidecomplete.pdf>
- The Foundation Center. (2013). Key Facts on U.S. Foundations: 2013 Edition. Retrieved from <http://Foundationcenter.org/gainknowledge/research/keyfacts2013/>
- Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. *New Directions for Evaluation*, 2009(123), 13–32.
- Wholey, J. S. (1979). *Evaluation--Promise and performance*. Washington, D.C.: The Urban Institute Press.
- Yin, R. K. (2009). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publications, Inc.