An Investigation of the Relationship Between Involvement In and Use of Evaluation in Three Multi-site Evaluations When Those Involved Are Not the Primary Intended Users

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Denise L. Roseland

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jean A. King and Frances Lawrenz, Co-Advisers

March 2011

ACKNOWLEDGEMENTS

I consider this work to be the culmination of efforts that moved me along a continuum that ranged from fear to faith. My first official day in the program, I put on a necklace with a simple pendant that was etched with the word “Believe”. That morning I hoped it would be enough to quell the irrational fear I felt about starting a Ph. D. program. I will take the necklace off for the first time when I successfully defend this work.

Along that same fear and faith continuum, many people have nudged, pulled, and prodded me closer to faith and farther from fear. First, to my former colleagues at the Wisconsin Department of Public Instruction, Dr. Margaret Ellibee and Barbara Bitters, you always said the right thing at the right time when I first began thinking about pursuing this degree.

So many of my evaluation studies classmates challenged my thinking and supported my development. I hope you know how much I valued the time we shared in this process.

To my committee members, Dr. Melissa Anderson and Dr. Darwin Hendel and my advisers, Dr. Jean King and Dr. Frances Lawrenz, who made it possible for me to work on my dissertation many miles from the Twin Cities—Your willingness to advise via email and at unusual times of day allowed me to feel supported in this process even though I no longer had regular face-to-face access to colleagues or faculty. I have particular gratitude for the way in which Drs. King and Lawrenz took turns playing two

important roles: responding with amazing speed to my questions and saying just the right thing to keep me focused on moving ahead.

To my colleagues on the Beyond Evaluation Use research project, Dr. Lija Greenseid, Dr. Stacie Toal, Dr. Boris Volkov, Kelli Johnson, Gina Johnson, and Patti Ross—Thanks for helping me redefine myself as a scholar and an evaluator and for giving me plenty of reasons to laugh as we worked.

Next, to my family—I whole-heartedly appreciate all the times you "overlooked" my absence or excused my distractedness while I worked through courses and wrote this dissertation. I know none of you understand my new line of work but your enthusiasm and interest in what I do is appreciated. I love you all.

To my dearest friends, Brenda, Michelle, and especially Melissa—Your support was crucial and our regularly scheduled get-togethers made it possible for me to take myself and this process less seriously.

Most importantly, to my niece, Quinn, who is now 8 years old—May you have more faith than fear and may you always believe you can…

ABSTRACT

This dissertation research describes an investigation that explores the nature of the relationship between participation in evaluation and the use of evaluation findings and processes within three large-scale multisite evaluations. The purpose of this study is to test whether assumptions and theories about participation translate into evaluation use in the same ways as seen in single evaluation sites. Using canonical correlation analysis and a collection of 20 interviews, this study describes and tests the relationship between these two critical conceptual powerhouses in evaluation. Using data that were collected as a part of the NSF-funded research Beyond Evaluation Use (Lawrenz & King, 2009), this study found that some theories and beliefs about participatory evaluation contribute to use and influence in similar ways as single-site evaluations. The differences identified in this research highlight potential planning and implementation considerations that might allow multi-site evaluators and funders of multi-site evaluation to enhance use and influence of multi-site evaluations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

As far back as the early 1970s, evaluators have been curious about what happens when potential users interact with evaluation findings. Thus, it is not surprising that the use of evaluation processes and findings has been a major focus of empirical research within the field of evaluation since evaluations became widespread with the proliferation of Great Society programs in the mid-1960s. Evaluation, with its purpose of providing information for decision making, may be seen by some as a wasted expenditure if the evaluation process and findings have no impact on policies, on programs, or even on society at large. Just as program managers want to know if their activities are making a difference, the evaluation community has spent considerable time and energy collecting data on whether evaluation activities are worthwhile.

Statement of the Problem

Overview of Evaluation Use

Evaluation use, in its early days, was defined as the direct, immediate use of evaluation results in decision making and was labeled as instrumental use (Weiss, 1998). While this idea provided guidance to the field, evaluators realized that clients frequently ignored evaluation findings. The frequency of unused or underused evaluations became what Patton (1997) called a "utilization crisis." In response, evaluation researchers began studying evaluation use from many angles. Much of this next generation of research on use uncovered that stakeholders used evaluation, but not in ways that aligned with evaluators' expectations or recommendations. Weiss' empirical study of evaluation use

(1980) found that people charged with decision making rarely used evaluation results to answer or inform direct questions. Rather, participants in her evaluation use research reported changes in thinking, ideas, or attitudes. Weiss labeled this enlightenment, also commonly referred to as conceptual use (Weiss, 1998). A third general type of use emerged as yet another way to describe evaluation use: persuasive or symbolic use. This type of use is signified by evaluation results used as a way to legitimize or validate a program often in a context where program decisions are politicized. Finally, researchers have identified ways in which people use what they learn from participating in the evaluation process. Process use (Patton, 2008) is defined as "individual changes in thinking and behavior, and program or organizational changes in procedures and culture that occur during the evaluation process" (p. 155). This newly termed evaluation use seemed to indicate that the process of completing evaluative activities and using evaluative thinking actually helped evaluation stakeholders acquire new information, change behavior and learn new skills (Patton, 2008; Weiss, 2005; Alkin & Taut, 2003).

Over time, as the field of evaluation matured, inquiries into utilization have broadened, and additional distinctions have emerged. As a result, theoretical activity has reconceptualized the field's understanding of use. Recently, the term evaluation influence has been used when discussing broader, longer-term, more indirect influences of evaluation processes and products. Scholars now note that focusing solely on the direct use of either evaluation results or processes has not adequately captured broader-level influences (Kirkhart, 2000; Alkin & Taut, 2003; Henry & Mark, 2003; Mark & Henry, 2004, Christie, 2007).

Overview of Participatory Evaluation

Based on the assertion that use is enhanced when stakeholders participate, a number of participatory methods and theories about evaluation emerged in the 1990s. This notion of participatory evaluation has come to encompass all approaches where "members of the evaluation community and members of other stakeholder groups relative to the evaluand each participate in some or all of the shaping and/or technical activities required to produce evaluation knowledge leading to judgments of merit and worth and support for program decision-making" (Cousins, 2003). Practices that are built on these ideals of participatory evaluation may be labeled in many ways, including fourth generation evaluation (Guba and Lincoln, 1989), participatory evaluation (Cousins & Earl, 1992; Cousins & Whitmore, 1998), democratic deliberative evaluation (House & Howe, 1999), empowerment evaluation (Fetterman, 2000) and values-engaged Evaluation (Greene, 2005).

Participation and use in multisite evaluations

Discussion of multisite evaluations emerged in the evaluation literature during the late 1980's. Initially, Sinacore and Turpin (1991) offered only an implicit definition for multisite evaluations and discussed organizational and methodological issues unique to the multisite evaluation context. Implicit in their discussion was the idea that multisite evaluations involved multiple program sites and required the conduct of evaluation activities across the sites.

One of the early promises of multisite evaluation was that it provided opportunity for evaluators to demonstrate a level of methodological rigor with regards to

generalizability that was difficult to achieve in a single, local program evaluation. By its very nature, the collection of data from multiple sites promised to aid in determining the extent to which a program influenced different types of individuals in a way that offered more statistical power. As such, early multisite evaluations were not designed to attend to many principles of participatory evaluation. In turn, the same issue about the use of processes and findings of these evaluations concerned those who expected programs and policy makers to benefit from the process and products of evaluation in their work as decision makers.

Research Questions

This thesis will describe the nature of the relationship between participation in evaluation and the use of evaluation findings and processes within four large-scale multisite evaluations. The purpose of this study is to test whether assumptions and theories about participation translate into evaluation use in the same ways as seen in single evaluation sites. Using a variety of reasoned statistical and theoretical methods, this paper seeks to describe and test the relationship between these two critical conceptual powerhouses in evaluation. The following research questions guided this study:

Primary Research Question: What is the nature of the relationship between participation in and use of evaluation in a multisite evaluation setting?

Supporting Question 1: To what extent can evaluation use be explained by participation?

Supporting Question 2: What contribution does any single factor related to participation in evaluation make to the explanatory power of participation?

Supporting Question 3: Do certain factors related to participation in evaluation contribute to explaining the composite of factors that make up the idea of use?

Significance of the Study

This study contributes to the knowledge base of participation in and the use and influence of program evaluation in theory and in practice. First, in theory, this study was concerned with the effects of involvement in and use of evaluation in large scale, multi-site evaluations and particularly among those at the local project level—who are not the primary intended user of evaluation processes or findings. This change in the unit of analysis is important because the discrete findings and theories that have emerged from single sites have never been tested in a multi-site context, and the interaction of the complex network of a multi-site setting may produce differing effects. Understanding these potential differing effects of a multi site context is important because of the substantial amount of public funds used to support both the projects and the related evaluation efforts. Given that the notion of affecting the field is inherent in the dissemination requirements in the three targeted evaluations, it would serve the theory in the field if we understand more about the interaction of involvement and use in these multi-site settings.

Secondly, in practice, the current study focused on the participatory frameworks set out by Cousins and Whitmore (1998) and examined the relationship between participatory evaluation and evaluation use in large, multi-site evaluations as measured by Toal (2007) and Johnson (2008). By using the results from two of the only empirical inquiries related to multisite evaluation (Toal, 2007; Johnson, 2008) and examining the

relationship between the two, this research hopes to move knowledge in the field from that of descriptive characteristics to knowledge that facilitates direct activities that promote participation resulting in use. This is important because the number of large, multi-site programs in education has increased. As demands for these multi-site evaluations increase, so, too, does the need for program evaluations that accommodate larger, more dispersed projects that involve many local stakeholders in efforts that advance the overall program goals. These evaluations can still leverage the benefits of stakeholder participation to improve the quality and cost-effectiveness of the evaluations, as well as to enhance people's knowledge and use of the evaluation process and findings.

Limitations

This study is confined to describing and analyzing involvement in and use of evaluation by secondary (or "unintended") users of evaluation findings and processes in only three NSF multi-site program evaluations. More specifically, some unique factors of the NSF context may confine generalization of these findings to multi-site evaluations from other contexts. For instance, within NSF, these three programs emerged during a time where evaluations tracked institutional changes and associated changes in student achievement, teacher content knowledge, or other goals. These comprehensive evaluations were broad and highlighted the importance of collecting similar evaluation data across project sites and as such, made the NSF the intended user of the evaluation findings. This study will however, examine use by local projects funded under these three NSF programs, making them secondary users of evaluation processes or findings. Other

multi-site evaluation efforts may place different priorities on the types of data collected across sites or have differing expectations of funded projects with regards to evaluation.

Definition of Key Terms

A number of key terms form the foundation this study. They are defined here:

Program: A major centralized national funding effort.

Project: One of many smaller efforts funded under a single program.

Involvement: Active engagement in at least one phase of an evaluation (i.e., planning, implementing, or applying the results).

Evaluation use: The application of evaluation processes, products, or findings to directly produce an effect. A person uses evaluation processes, products, or findings.

Evaluation influence: The capacity of evaluation processes, products, or findings to indirectly produce a change in understanding or knowledge. An evaluation process, product, or finding influences a person or society.

Overview of the Dissertation

Chapters 2 through 5 comprise the remainder of this dissertation. Chapter 2 reviews the pertinent literature related to participatory evaluation, evaluation use, and multi-site evaluations and describes the empirical or theoretical nature of the research on these constructs. Chapter 3 describes the methodology of the dissertation. Chapter 4

presents the results. Chapter 5 is a discussion of the findings and concludes with implications of this research.

CHAPTER 2

LITERATURE REVIEW

Introduction

This chapter provides a review of the literature related to participatory evaluation and the use of evaluation processes and products in a multisite evaluation context. While the literature pertaining to the relation between participation in and use of evaluation is extensive, little has been done to date to examine the interrelationship of the ideas in a multisite evaluation context. This review will provide an overview of the history of research about participatory evaluation, evaluation use, and prevailing thought about the uniqueness of multisite evaluation contexts, especially where the ideas of participation and use intersect.

Participatory Evaluation

While the idea of stakeholder involvement in evaluation became a major topic in the field of evaluation in the 1970s, the definition of participatory evaluation emerged as two parallel paths of thought or philosophical orientations of evaluation and appears to be rooted strongly in the ideas of action research (Whyte, 1990; Levin, 1993). For example, Levin (1993) posited that research knowledge and evaluation data are valid only when informed by the perspectives of practitioners. This philosophical orientation is evident in the pattern of thought that emerged regarding stakeholder-based evaluation where beliefs focused on the inclusion of those who make decisions about a program or those whose lives are affected by the program and its evaluation (Cousins & Whitmore, 1998). Involving these groups meant involvement in decision making at all phases of an

evaluation. Along one path, evaluation theorists like House (1976) started writing about involving stakeholders in evaluation as a way of giving voice to the powerless or disadvantaged, and as such, serve a social justice role. Along the other path, the role of stakeholder involvement in evaluation as a means to cultivate evaluation utilization[1] emerged from thought leaders like Alkin (Alkin et al., 1979), Patton (Patton et al., 1977), and Rich (1977). Evaluation theory has since advanced to explicitly describe several differing participatory approaches to evaluation. More specifically, approaches that stress social justice concerns include fourth generation evaluation (Guba & Lincoln, 1989), empowerment evaluation (Fetterman, 1994), transformative evaluation (Mertens, 1999), and deliberative democratic evaluation (House & Howe, 2000). In addition, there are several explicit approaches concerned with evaluation utilization that include responsive evaluation (Stake, 1975), utilization-focused evaluation (Patton, 1978), and participatory evaluation (Cousins & Earl, 1992; Cousins & Whitmore, 1999). It is not the purpose of this review to explicate the distinctions between and among these approaches other than to note the following: 1) differences do exist that may have implications for studying the process of stakeholder involvement, and 2) the existing empirical base for participation in evaluation suggests that active involvement does lead to increased levels of use (Cousins, 2003).

Regardless of the pragmatic or emancipatory orientation, participatory evaluation represented a shift in the consciousness of some in the evaluation profession from being a profession of dispassionate outsiders to one where evaluators became a facilitator and

[1] For the sake of this research, the terms evaluation use and utilization will be used interchangeably to mean the application of evaluation processes, products, or findings to produce an effect.

coach alongside program staff. Burke (1998) notes that all participatory evaluation methodology shares a common set of principles and a process for engagement with stakeholders. However, Cousins (2003) cautions that "participatory evaluation" has come to represent all collaborative or stakeholder-based approaches to evaluation and may result in confusion about conditions for or consequences of use of these approaches in various contexts. Before looking at the empirical literature regarding participatory evaluation and its relationship to use, it is important to review the theoretical frameworks of Cousins and Whitmore (1998) and Cousins (2003).

Cousins and Whitmore (1998) began their discussion by delineating what they interpret, from a robust review of the literature, to be two primary forms of participatory evaluation: practical and transformative. Practical participatory evaluation (P-PE) serves a central function of fostering evaluation use and thus supports program or organizational decision making or problem solving. Transformative participatory evaluation (T-PE) serves to empower members of community groups who are less powerful to engage in inquiry in ways that are authentic to the groups involved. While this distinction highlights the disparate nature of participatory practice, it is worth noting that the line that divides these concepts is not an impermeable boundary. Rather, a P-PE project may empower participants while maintaining its focus on the use of evaluation processes and findings. Likewise, a T-PE project may generate practical value for those who operate a given program or benefit from its work.

Cousins and Whitmore (1998) also offer a second framework that provides a set of process dimensions that allow researchers and practitioners alike to distinguish forms

of systematic inquiry within collaborative forms of evaluation. This framework seems to be based in a single-site context and makes no explicit connection to dimensions that may exist in a multisite setting. Figure 1 shows Cousins and Whitmore's framework as a three-dimensional plus sign, with each plane of the image representing one of the dimensions of the conceptual framework. The three process dimensions Cousins and Whitmore propose are: 1) control of the evaluation process; 2) stakeholder selection for participation; and 3) depth of participation. For each process dimension represented in this framework, there is a continuum for the characteristic described. For instance, the dimension dealing with control of the evaluation (plane labeled as A in the diagram) shows the opposing ends of the continuum as evaluator controlled and practitioner controlled and relates particularly to technical decisions regarding evaluation processes and conduct. Depth of participation (plane labeled as C in the diagram) ranges from deep participation of stakeholders in all aspects of the evaluation to consultation where stakeholders have no decision-making control or responsibility. Finally, stakeholder selection for participation (plane labeled as B in the diagram) is bound by all legitimate groups on one end to primary users on the other end.

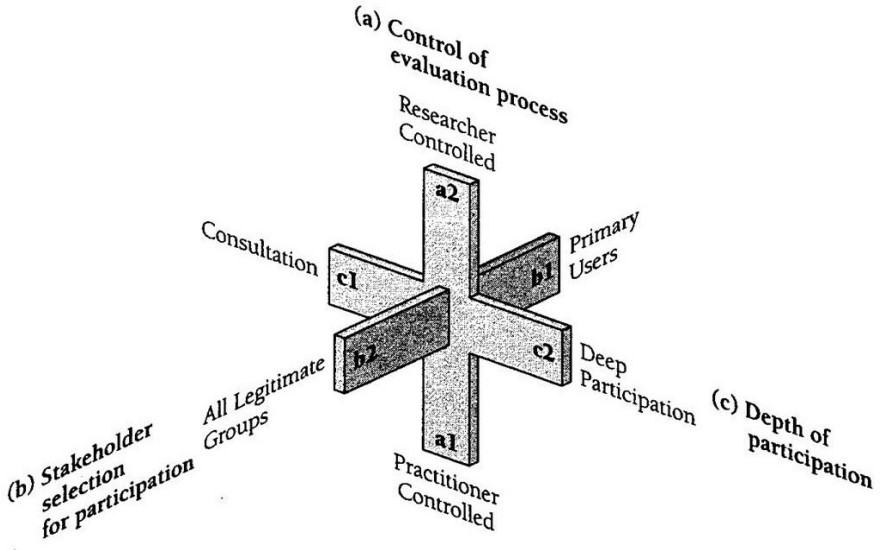


Figure 1. Cousins and Whitmore's dimensions of collaborative inquiry

In their examination of participatory practices in multi-site evaluations, Lawrenz and Huffman (2003) also use the dimensions to categorize the participatory nature of multisite evaluations. The Cousins and Whitmore (1998) framework examines dimensions of participatory evaluation in such a way that is useful for this dissertation research and will thus serve as the conceptual framework for the review of the literature regarding participatory evaluation that follows.

Control of the Evaluation

The first distinguishing characteristic of the Cousins and Whitmore (1998) model deals with control of the evaluation process, ranging from control of decisions being completely in the hands of the evaluator to control residing completely with the program practitioner. This idea of decision-making involvement or control emerged in the work of Greene (1987) and Mark and Shotland (1985) and prompted evaluators to think about the roles of program participants and evaluators when beginning an evaluation project.

Cousins and Whitmore make the distinction that "control" as used in this characteristic of the framework relates to technical decisions about evaluation processes and conduct. King and Stevahn (2002) also highlight a similar continuum of control and note that the relationship between evaluators and program participants differs. In Figure 2, King and Stevahn illustrate this idea of balance of control by plotting evaluator and program participants on separate axes that show how the role of the evaluator changes as program participants assume more or less control over evaluation decision making. The diagram shows that, at one end of the figure, evaluators design and implement a study; at the other, clients conduct a self study from beginning to end with support from the evaluator. Along the continuum are various intermediate participatory and collaborative possibilities where evaluators and clients serve as collaborators during the design and implementation processes. King and Stevahn also note that only the collaborative and participant-directed ranges (the two on the right side of the figure) of the model represent participatory approaches to evaluation.

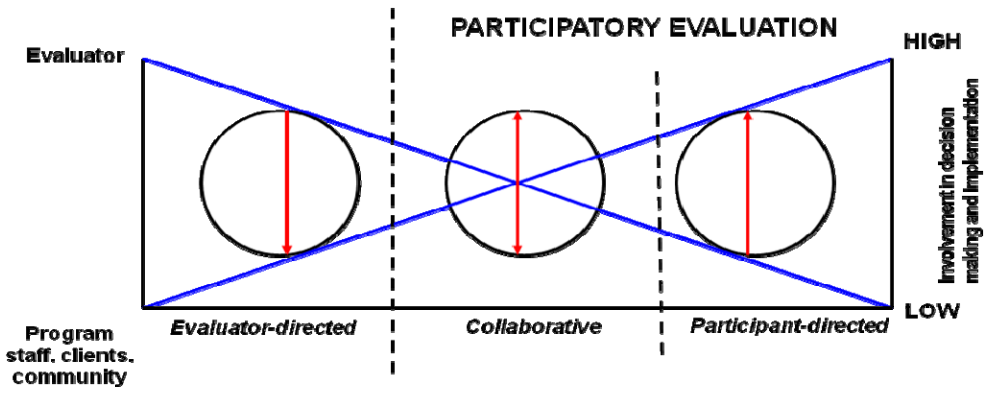


Figure 2. The Interactive Evaluation Practice Continuum (King & Stevahn, 2002)

Similarly, Themessl-Huber and Grutsch (2003) examine the locus of control in participatory evaluation. They assert that "the person whose decisions establish the course of the evaluation" (p. 93) differs depending on the phase of the evaluation. While they accept Cousins and Whitmore's (1998) framework, they believe that the shifting locus of control results from the evaluator's responsiveness to the stakeholders' needs rather than aligning statically with a single point on the continuum as might be inferred from Cousins and Whitmore's work.

Stakeholder Selection

The second dimension of Cousins and Whitmore's (1998) framework is stakeholder selection, i.e., deciding who participates actively in the evaluation. At one end of Cousins and Whitmore's framework, we see that the inclusion of stakeholders is limited only to potential users of the evaluation processes and products and, on the other end, the inclusion of stakeholders is open to all legitimate groups. Thinking back to the other framework Cousins and Whitmore (1998) reference in this work, we might expect

transformative participatory evaluation approaches to be inclusive of more stakeholder groups than the practical participatory approaches.

Participatory evaluation is rooted in the idea that multiple stakeholders play an active role in the evaluation. The range represented in Cousins and Whitmore's framework highlights the differing views of how "active" is operationalized in different evaluations. Many evaluators (Gold, 1983; Mark & Shotland, 1985; Mercier, 1997; Greene, 2000; King & Ehlert, 2008) note the logistical, interpersonal, cultural, and financial challenges to involving a truly diverse set of stakeholders at all stages of the evaluation process. As a result, the inclusion of diverse stakeholders is one aspect of participatory evaluation that splits in the practice of evaluation between involving a few stakeholders in depth and involving a broad range of stakeholder groups less actively (Taut, 2008). This section will conclude with examples from the literature of practical considerations evaluators face in making the decisions surrounding stakeholder involvement while further discussion of the depth of participation follows in the next section.

Mark and Shotland (1985) discuss the role that power and legitimacy play in the selection of stakeholders to involve in evaluation. An evaluator who is concerned with evaluation use would tend to select high-powered stakeholders who have the authority to make decisions and act on the findings. Similarly, an evaluator who desires social justice through evaluation would be most likely to favor involving low-powered stakeholders as a means to give that group voice. This value judgment of the evaluator, the focus on utilization versus social justice, is a factor that affects the selection of stakeholders and

comes with ethical implications for evaluators. For instance, evaluators grapple with the trade-offs among adequate representation, technical accuracy, and feasibility as they determine which stakeholders will be involved. A narrow definition risks missing important groups and individuals whose exclusion may jeopardize the usefulness and validity of evaluation findings. A wider, more inclusive definition of appropriate stakeholders may jeopardize the feasibility of the evaluation itself, by increasing the complexity and costs of managing a large and diverse stakeholder group.

Wallace (2008) describes, in a case, that the stakeholders chosen to participate in evaluation might best be informed by the stages of the evaluation, i.e., involving a diverse group of stakeholders during the creative process of designing the evaluation, but narrowing the stakeholder groups involved during data collection and analysis. Similarly, King and Ehlert (2008) discuss the benefits of involving fewer stakeholder groups more deeply throughout the entire evaluation process while involving a larger number of stakeholders at the beginning and end of the evaluation process. This rationale for selecting stakeholders for involvement allowed the evaluators to minimize the impact of complicated logistical arrangements and conflicts that may arise with multiple stakeholder groups involved and responsibly use financial resources allocated to the evaluation.

Greene (2000) describes challenges faced in an evaluation that was explicitly designed to promote democratic dialogue among all legitimate stakeholder groups. First, she notes that absence of participation by stakeholder groups, despite evaluator vigilance inviting the groups to participate. Second, methodological discussions often superseded

conversations about the evaluation's results, i.e., the value and meaning of the data. Finally, Greene describes how the client selected stakeholder participants with little heed to her warnings to involve all relevant and often overlooked stakeholders. Eventually, the stakeholder group was dominated by program opponents, resulting in an evaluation that was of limited use to the program and threatened the credibility of the evaluator and the evaluation process.

The previous examples highlight the more inclusive range of stakeholder selection described by Cousins and Whitmore (1998). The other end of the continuum, however, calls only for involvement of primary intended users, limiting those selected for involvement to be those with the requisite knowledge and ability to the design of the evaluation. Patton (2008) explains that no evaluation can address all stakeholder needs or demands equally well. As a result, Patton recommends a stakeholder selection process that limits the involved stakeholders to those who are in a position to make decisions regarding the program under examination and act on the findings of the evaluation. Within Patton's view, the identification of stakeholders chosen to participate is subject to negotiation with primary intended users and is contingent on the evaluation's purpose and other situational factors, which the evaluator should be aware of at all times during the evaluation.

Depth of Participation

The third dimension of Cousins and Whitmore's (1998) framework is depth of participation. This continuum ranges on one end from consultation with no decision-making control or responsibility to deep participation that involves stakeholders in all

aspects of the evaluation including design, data collection, analysis, and reporting and dissemination. The empirical research on this dimension of the framework offers limited information about the ways in which stakeholders are involved. For example, a 1996 survey of Canadian and American evaluators suggests that stakeholder involvement is most often limited to evaluation tasks that are "not heavily technical" (Cousins et al., 1996). A similar survey conducted in 2006 showed that evaluators most often involved stakeholders only when focusing the evaluation (Fleischer & Christie, 2009). Cousins (2003) points to a key challenge related to the depth of stakeholder involvement in participatory evaluation: the need for technical sophistication and external credibility of the evaluation, on one hand, and the responsiveness and extent of stakeholder participation, on the other hand. He points out that the active involvement of stakeholders can introduce bias and give advantage to a select group. In addition, when an evaluator serves as facilitator who supports program participants who play an active role in the evaluation, it can also be perceived that the evaluator loses independence by giving up control of the process (Stufflebeam, 1994).

Burke (1998) proposes several key decision points that represent opportunities for meaningful stakeholder participation. Burke writes that the participatory nature of an evaluation can be seen as the extent by which stakeholders participate in the following decision points:

1. Deciding to do the evaluation;
2. Assembling the team;
3. Making the evaluation plan;

4. Collecting the data;

5. Synthesis, analysis and verification of the data;

6. Developing future action plans; and

7. Dissemination and use of the outcomes of evaluation activities

Examples of these key decision points and the varying ways in which they are operationalized by evaluators are evident in the literature. Ayers (1987) describes how stakeholders in an evaluation created instruments, collected data, conducted data analysis and wrote and disseminated findings—placing this example on the deep participation end of the continuum. Similarly, Torres et al., (2000) describe an evaluation where stakeholders contributed significantly to the overall evaluation design and implementation. In contrast, Brandon (1998) describes an evaluation where stakeholders' depth of participation was limited to helping the evaluators understand more about the program and reviewing evaluator recommendations—a role much more oriented towards one-sided information sharing and verification of evaluator's ideas. Brandon still considers this to be participatory in nature because the evaluation used carefully developed methods for involving stakeholders that ensured no group's expertise was ignored.

While Cousins and Whitmore (1998) call on all of us to consider that the three dimensions of their framework may be independent of one another, Taut (2008) points out that some scholars (Alkin et al., 1998; Greene, 2000; House, 2003; Mark & Shotland, 1985) posit that the decisions regarding the selection of stakeholders and the prioritization of the breadth/depth of their involvement must be made by the evaluator

"based on the values he or she holds related to the main purpose or role of evaluation" (p. 225). Still, Taut points out that regardless of the evaluator's intention to be inclusive and foster participation, "…few evaluations have the resources necessary to involve the full range of stakeholders in a just and equitable manner as well as deeply and actively. In other words, feasibility usually demands a tradeoff between depth and breadth of stakeholder participation" (Taut, 2008, p. 225). Information in Table 1 below helps delineate the empirical and theoretical literature on participatory evaluation discussed thus far.

Table 1.

Summary of Participatory Literature Discussed

Author(s) & Year	Theoretical or Empirical nature	Design, Methods, or Scope	Description	How the work aligns with Cousins & Whitmore's three process dimensions
Ayers (1987)	Empirical	Case study	Describes a single evaluation where stakeholders assume and active role in all phases of the evaluation	Depth of participation
Brandon (1998)	Theoretical	Case study	Validity of evaluation claims can be improved through stakeholder participation	Selection of stakeholders
Burke (1998)	Theoretical	Reflective case narrative	Describes stakeholders involved in stages of decision making throughout the evaluation process	Control of evaluation
Gold (1983)	Theoretical	Reflective case narrative	Degree of evaluator effectiveness in using strategies that involve stakeholders contributes to strength or limitation of evaluation outcomes	Depth of participation
Greene (2000)	Empirical	Case study	Describes the challenges of an evaluation that deliberately sought to involve a broad array of stakeholders	Stakeholder selection and depth of participation
King & Ehlert (2008)	Empirical	Reflective case narrative	Describes the challenges faced when trying to involve stakeholders in evaluative activities	Control of evaluation and depth of participation
King & Stevahn (2002)	Theoretical		Discusses the varying role of the evaluator in a participatory evaluation context	Control of evaluation and depth of participation

(cont.)

Author(s) & Year	Theoretical or Empirical nature	Design, Methods, or Scope	Description	How the work aligns with Cousins & Whitmore's three process dimensions
(Table 1. cont.)				
Lawrenz & Huffman (2003)	Theoretical	Examines data and literature from five National Science Foundation education programs	Proposes a model for implementing multisite participatory evaluations	Depth of participation
Mark & Shotland (1985)	Theoretical		Discusses the role of evaluator values in the selection of stakeholders and the consequences of stakeholder participation	Stakeholder selection
Taut (2008)	Theoretical	Literature review	Discusses the distinction between deep involvement of a few stakeholders versus active involvement of a broad array of stakeholders	Depth of participation
Themmesl-Huber & Grutsch (2003)	Empirical	Case study	Discuss the changing nature of the locus of control during different phases of an evaluation	Control of evaluation
Torres et al., (2000)	Empirical	Reflective Case Narrative	Summarizes benefits to stakeholders of their collaborative evaluation mainly that time for reflection and sense-making provided advantage for the most involved stakeholders	Depth of participation
Turnbull (1999)	Empirical	Structural equation modeling analysis	Tested a series of causal relations in a proposed model of participatory evaluation in order to better understand participatory evaluation	All
Wallace (2008)	Empirical	Case study	Addresses the tradeoff evaluators face when involving stakeholders in the entire evaluation process	Depth of participation

While the Cousins and Whitmore framework has shaped thought regarding participatory evaluation for decades, a paucity of empirical research on participatory evaluation leaves evaluation scholars still wondering about many aspects: appropriate methods for indentifying stakeholders to involve, specific examples of how and when to involve stakeholders, and processes or mechanisms to effectively engage stakeholders in different aspects of evaluation tasks. This dissertation research will address some of these weaknesses in the current literature as it seeks to identify the aspects of participation that result in use in a multisite setting.

Evaluation Use

Christie (2007, p. 8) notes, "Evaluation utilization is arguably the most researched area of evaluation and it also receives substantial attention in the theoretical literature." It is not surprising that the use of evaluation processes and findings has been a major focus of empirical research within the field of evaluation. Evaluation, with its purpose of providing information for decision making, is seen as a wasted expenditure if evaluation processes and findings have no impact on policies, on programs, or, as an extension, on society at large. Thus, since evaluations became widespread with the proliferation of Great Society programs in the mid-1960s, identifying the types of evaluation use has been the subject of continued research and discussion.

Before discussing the four standard conceptualizations of evaluation use that emerge in the literature, it is worth noting that the early research on evaluation use found that evaluations were not being used in government decision making on policy matters (Weiss, 1972, 1990). This lack of use drew considerable concern at a time as another

corresponding trend showed a notable increase in funding designated for evaluation of program or policy initiatives. This intersection of trends prompted Patton (2008) to declare a "utilization crisis." Subsequent research showed an increase in evaluation use, but "use" needed to be broadly defined and was best unearthed using naturalistic methods. The research also indicated that the idea of program planners waiting for formal evaluation findings and then directly using the results to make changes to a program was a myth. For instance, King and Pechman (1984) describe "charged use" in a longitudinal case study as a form of use that occurred when program participants interacted with the evaluators and the evaluation process that had the potential for disruption or change in subsequent actions of the program staff. Preskill and Caracelli (1997) refer to the "cognitive and behavioral changes resulting from users' engagement in the evaluation process." Weiss (1998) expands the discussion on change that results from evaluation by surmising that the pathways to change in individual practices are likely to differ greatly from the pathways to collective change including policy change.

Types of Evaluation Use

After decades of research on evaluation utilization, Alkin and Taut (2003) succinctly describe two distinct aspects of use: process use and use of evaluation findings. The use of findings is traditionally divided into three types: instrumental, conceptual, and symbolic (Leviton & Hughes, 1981). Instrumental use refers to instances where someone has used evaluation knowledge directly. Conceptual use refers to cases when no direct action has been taken, but where people's understanding has been affected. Symbolic use refers to examples when a person uses the mere existence of the

evaluation, rather than any aspect of its results, to persuade or convince. Process use is the newer concept, defined by Patton (2008, p. 155), as "individual changes in thinking, attitudes, and behavior, and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process." Amo and Cousins (2007) point out that research on this type of evaluation use is in its infancy. Importantly, unlike the use of results, process use is more under the control of the evaluator in that an evaluator can more easily make changes to the evaluation process with the hope of facilitating use. While there can be overlap between process use and instrumental and conceptual uses, one notable distinction is that process use does not require programmatic changes to occur after the delivery of findings (Shulha & Cousins, 1996). Table 2 below defines these four conceptualizations of use in more detail and provides general examples.

Table 2.

Types of Evaluation Use

Type of Use	Definition	Examples
Instrumental use	Using evaluation findings as the basis for action	Eliminating an ineffective program Making changes to a program based on evaluation findings Allocating resources for a program differently based on evaluation findings
Conceptual use	When an evaluation influences decision makers' thinking about a program	Becoming aware of evaluation results Becoming aware of formerly implicit program features and mechanisms after learning about evaluation results Developing opinions, attitudes and knowledge about evaluation in general
Symbolic use	Use of evaluation findings for political self-interest	Use of evaluations results to justify decisions already made about a program Use of evaluations to advocate issues and persuade people to act
Process use	Behavioral and cognitive changes in persons involved in evaluations as a result of their participation in evaluation activities	Learning to think evaluatively while completing evaluation tasks (i.e., developing a logic model, identifying evaluation priorities, participating in evaluation design) The observance of instrumentation effects (the idea that "what gets measured gets done")

Adapted from Johnson (1998) and Patton (2008)

The Evolution of Thinking on Use

Over time, as the field of evaluation has matured, inquiries into utilization have broadened, and additional distinctions have emerged. As a result, theoretical activity has reconceptualized the field's understanding of use. Recently, the term evaluation influence has been used when discussing broader, longer-term, more indirect influences of

evaluation processes and products. Scholars now note that focusing solely on the direct use of either evaluation results or processes has not adequately captured broader-level influences (Kirkhart, 2000; Alkin & Taut, 2003; Henry & Mark, 2003; Mark & Henry, 2004, Christie, 2007). Kirkhart (2000) advocates that the terms use and utilization in the field of evaluation should be abandoned in favor of using evaluation influence as a unifying construct. Kirkhart's "integrated theory" defines influence as "the capacity or power of persons or things to produce effects on others by intangible or indirect means" (2000, p. 7). Kirkhart envisions three dimensions of evaluation influence, represented as a cube-like figure: source (the active agent of change or evaluation process or results), intention (the extent to which there is intended or unintended influence exerted), and time (immediate, end-of-cycle, long-term) seen in Figure 3 below. She notes that while these dimensional subdivisions are somewhat arbitrary, they would better capture effects that are incremental, unintentional, and multidirectional alongside the intended, instrumental, and episodic effects already well represented in the well researched ideas of use.

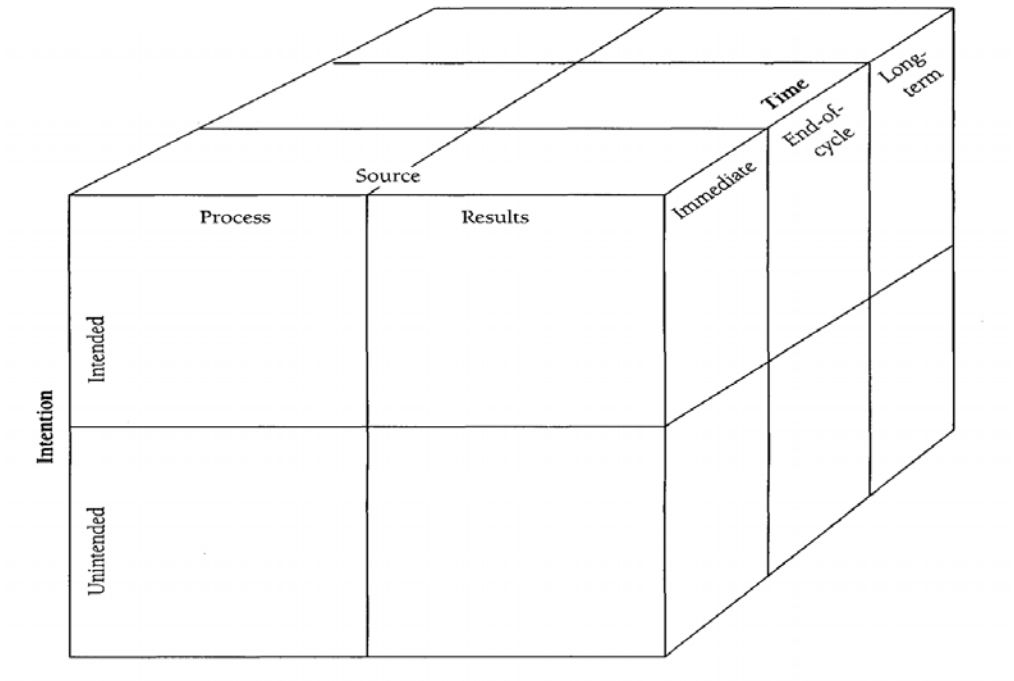


Figure 3. Kirkhart's integrated theory of influence

Mark and Henry (Mark, Henry, & Julnes, 1999; Henry & Mark, 2003; Mark & Henry, 2004) also push for broadening the way evaluators conceptualize the consequences of their work. They argue that the goal of evaluation is social betterment and suggest the need to identify the mechanisms through which evaluations lead to this ultimate goal along differing paths of influence and at different levels (i.e., individual, interpersonal, and collective). Mark and Henry map out a logic model for evaluation, focusing on evaluation consequences related to the improvement of social conditions. Just as program theory connects program activities with outcomes while also explaining the processes through which the outcomes are achieved, Henry and Mark's program theory of evaluation identifies evaluation as an intervention with social betterment as its ultimate outcome (Johnson et al., 2009). Mark and Henry label traditional notions of

instrumental, conceptual, and persuasive use more specifically as, for example, skill acquisition, persuasion, or standard setting. These, then, become the mechanisms through which social betterment is achieved.

Several have tried to study the idea of influence. Long before Kirkhart's model, Weiss (1980, 1982) examined evaluation influence on decision making and found that decision makers were more likely to pay attention to results that confirm what they already believe. Christie (2007) conducted a simulation study to examine evaluation influence at the first level of the Henry and Mark framework, the individual, in an attempt to provide a general understanding of the types of information that exert influence on program decisions made by decision makers. The results showed that evaluation information from large-scale evaluations and case study data have more influence on the decision-makers' program decisions. Christie posits that the influence of these forms of evaluation information resulted from the social and political context of the evaluation. Similarly, Roseland et al. (2009) examine the broad impact of four program evaluation projects on the field, rather than the uses and influence on the intended user. This study, which sought to measure something that had not previously been measured, included a variety of methods to measure a concept as broad as "influence." Findings indicated that the four programs examined had only limited influence on the field, but illuminated a number of factors that might improve future research on the topic, including use of social network analysis and the importance of using viable methods for identifying an appropriate sample of a somewhat amorphous population.

Building on the ideas of Mark and Henry, Alkin and Taut (2003) carefully distinguish between evaluation use and influence. To them, evaluation use "refers to the way in which an evaluation and information from the evaluation impacts the program that is being evaluated" (Alkin & Taut, 2003, p. 1). In their view, evaluators are aware of these evaluation impacts, both intended and unintended. By contrast, "the concept of influence adds to the concept of use in instances in which an evaluation has unaware/unintended impacts" (p. 9). The distinction of unaware/unintended impacts is of particular importance in this research as it seeks to better explain the relationship between involvement in and use of evaluation in a multisite context where dissemination of processes and findings are expected of local project staff in an effort to advance knowledge or skill in the field. This dissemination of processes, products, and findings to other program participants could lead to impact that a multisite evaluator is unaware of or did not plan.

Patton (2008) argues that the idea of influence calls attention to the importance of "looking for unintended effects; examining long-term, incremental and unanticipated uses of findings; and investigating diverse forms of influence" (p. 110). Similarly, Caracelli (2000) notes that Kirkhart's theory of influence provides "a framework that allows for the incorporation of different paradigms and the opportunity to advance dialogue by encouraging the inspection of language and meaning" (p. 109).

Most recently, Brandon and Singh (2009) set out to critique the research on evaluation use and discussed the balance of the types of methods and the implications of this balance for making conclusions about use. Brandon and Singh examine 52 empirical

studies specifically about evaluation use and provided useful descriptions of the literature. Their review of the literature notes that empirical studies on evaluation use provide much less information about the level of use than they do about the relationship of use with various characteristics of the evaluation. In addition, Brandon and Singh found that studies of evaluation use in the field of education were fairly well balanced across several methods: surveys, simulations, case studies, and "narrative reflections." If we assume that a criterion for the soundness of a body of studies is that they consist of a mix of quantitative and qualitative approaches and various kinds of study designs, then Brandon and Singh's findings adequately address the criterion related to balance of methods.

Regarding a second criterion—the implications of balanced methods for drawing conclusions about use—Brandon and Singh report that, with notable exceptions, the methods of the body of studies were unsound because they often lacked sufficient information about the methods used to study evaluation. This lack of sufficient information in the studies examined presents opportunity for numerous violations related to content validity, including a lack of sound technical quality of instruments and their administration along with the overreliance on self-reported (and potentially biased) data in the empirical literature on evaluation use. These findings might lead us to conclude that the studies included in their examination might not provide definitive evidence about the effects on use, but they provide suggestions that help evaluators improve the likelihood of use and point the way to improved research on evaluation use.

Table 3 differentiates between the empirical and theoretical literature on evaluation use discussed to this point. When scholars review the literature on evaluation use, a recurring criticism stands out: the empirical work that is available offers findings from a flawed standard of evidence. There is often an over-reliance on people's self-reported information about use (Brandon & Singh, 2009). In addition, these self-reported data are often not validated through other means nor is there systematic triangulation of information. This dissertation research, however, is designed with these limitations in mind. While self-reported data about use will be used, care will be taken to validate these findings with information from other methods of inquiry as well.

Table 3.

Summary of Literature on Use Discussed

Author(s) & Year	Theoretical or Empirical nature	Design/ Methods/ Scope	Description
Alkin & Taut (2003)	Theoretical	—	Disagrees with Kirkhart's assertion that all the dimensions of influence are equal. Alkin & Taut argue that intention is weighted differently
Kirkhart (2000)	Theoretical	—	Proposes three dimensions of influence: source, intentions, and time
Leviton & Hughes (1981)	Theoretical	—	Demarcates use, dispelling beliefs that it is a unitary concept, by describing the interaction of several dimensions (instrumental, conceptual, and symbolic).
Mark & Henry (2004); Henry & Mark (2003)	Theoretical	—	Calls on evaluators to expand their conceptions of use and consider influence (all consequences of evaluation, not just the intended consequences)
Shulha & Cousins (1996)	Theoretical	—	Discusses how both empirical and conceptual research on the nature, causes, and consequences of utilization has become immersed in issues of context
Amo & Cousins (2007)	Empirical	Bibliographic methods	Examines how empirical research has examined process use
Brandon & Singh (2009)	Empirical	Bibliographic methods	Assesses the strength of methodological warrants for the findings of empirical research on program evaluation use
Christie (2007)	Empirical	Simulation study	Not all evaluation information is equal in the eyes of decision makers' who use evaluation information to determine their potential actions
Greene (1987)	Empirical	Case study	Stakeholders report cognitive, affective, and political use
King & Pechman (1984)	Empirical	Narrative case reflection	Describes the results of a case study that examined the process of local evaluation use

(cont.)

Author(s) & Year	Theoretical or Empirical nature	Design/ Methods/ Scope	Description
(Table 3. cont.)			
Mercier (1997)	Empirical	Case study	In a multisite setting, despite involvement at all levels, there was little direct impact or use of evaluation process or findings
Preskill & Caracelli (1997)	Empirical	Survey: 530 North American TIG members (54% response rate)	Examines American Evaluation Association Evaluation Use Topical Interest Group (TIG) members' perceptions about and experiences with evaluation use
Roseland (2009)	Empirical	Survey: chi-squares and social network analysis	Discusses the challenges of measuring "influence" on an amorphous population like "the field"
Turnbull (1999)	Empirical	Survey: Structural equation modeling analysis	Examines the effects on symbolic and instrumental use that result from various characteristics of participatory evaluation
Weiss (1980, 1982)	Empirical	Case Study	Highlights that decision makers are more likely to pay attention to results that confirm what they already believe

The Intersection of Participation and Use

Other scholars have critically examined the literature on evaluation use. One seminal work by Cousins and Leithwood (1986) prepared a synthesis of 65 empirical research studies on evaluation use published between 1971 and 1985. Greenseid (2008) and Johnson et al., (2009) picked up where Cousins and Leithwood left off and, using the same analytic framework, reviewed the literature on use from 1985–2005. The field changed considerably in the 20 years since Cousins and Leithwood conducted their

review. Key concepts were introduced, including process use and evaluation influence, and more sophisticated qualitative, quantitative, and mixed-method methodologies were developed.

Similarly, Taut (2008) and Toal (2009) examined the literature on participatory evaluation. Taut notes that the literature on stakeholder involvement leaves much to be studied and understood related to stakeholder involvement in program evaluation. This is due to the limited empirical base regarding the contributions and limitations of stakeholder involvement in terms of various outcomes, making it challenging for us to understand the conditions under which stakeholder involvement does or does not work. She proposes much could be gained by conducting mixed method studies of stakeholder involvement in different settings. Toal et al., (2009) note that it is difficult to gain insight into participatory evaluation in multisite settings from findings based in small, localized program evaluations (as is the case with most research on participatory evaluation). The contextual differences of multisite settings prompted Toal et al., (2009) to advocate for the use of the term involvement instead of participation when examining multisite settings.

One thing is evident in existing reviews: much of the literature on participation and use has come from bounded environments, i.e., single-location local programs. This dissertation research, however, is concerned with the effects of participation in and use of evaluation in large scale, multisite evaluations. Inevitably, this changes the unit of analysis from discrete findings and actions to understanding how an entire network is affected by the infusion of evaluation processes and evaluation findings.

Multisite Evaluations

Discussion of multisite evaluations emerged in the evaluation literature during the late 1980's. Initially, Sinacore and Turpin (1991) offered only an implicit definition for multisite evaluations and discussed organizational and methodological issues unique to the multisite evaluation context. Implicit in their discussion was the idea that multisite evaluations involved multiple program sites and required the conduct of evaluation activities across the sites. Furthermore, Sinacore and Turpin (1991) described two distinguishing forms of multisite evaluations: 1) a program that is implemented in identical ways at different geographical locations, 2) a program that is implemented in different ways (yet unified by a shared goal) at different geographical locations. Both of these forms of multisite evaluation present operational and methodological challenges for evaluation practitioners, and some of these challenges have implications for participatory evaluation and evaluation use. This section will highlight the current understanding of multisite evaluation as it has evolved over the last 20 years and how evaluators have tried to address common challenges of the multisite context.

One of the early promises of multisite evaluation was that it provided opportunity for evaluators to demonstrate a level of methodological rigor with regards to generalizability that was difficult to achieve in a single, local program evaluation. By its very nature, the collection of data from multiple sites promised to aid in determining the extent to which a program influenced different types of individuals in a way that offered more statistical power. Sinacore and Turpin argued that the variation in participants that is created by using multiple sites allows the evaluator to examine the extent to which a

program influences different types of people in a much broader way than "deliberate sampling for heterogeneity" (1991, p. 8). Given the two forms of multisite evaluation, this promise of increased generalizability is threatened by differences across program sites when the second form of multisite evaluation—different implementation across sites—is implemented. Validity can be further threatened in a multisite context by the role played by stakeholders in data collection and analysis at the many individual program sites. Members of stakeholder groups may not possess the knowledge or skills to engage in data collection or analysis in ways that protect against data contamination or bias, as a professional evaluator presumably would.

Additional reasons for conducting multisite evaluations beyond obtaining an adequate sample size emerged in the literature (Barley & Jenness, 1993; Worthen & Schmitz, 1997) and included rapid accumulation of knowledge, increased policy impact of a particular funding initiative, increased rate or amount of systematic change, or opportunity for evaluators and other stakeholders to work together to examine difficult issues and build local evaluation capacity.

Sinacore and Turpin (1991) noted a number of administrative issues evaluators face in multisite evaluations including the development of staffing plans, training, and selection of site coordinators. These challenges have a direct relationship to ideas raised in the previous section on participatory evaluation because they illustrate the tradeoffs that Taut (2008) discusses regarding breadth versus depth of participation of stakeholders. In addition, Sinacore and Turpin (1991) highlight issues related to cross-site communication, including the nurturing of stakeholder involvement. The limited amount

of literature that discusses multisite evaluations tends to focus on these pragmatic challenges and a discussion of ways in which the evaluator attempted to overcome obstacles.

In the 10 years that followed Sinacore and Turpin's work, interest in multisite evaluations grew as federal agencies and foundations that provided programmatic monies encouraged the use of multisite evaluations. Straw and Herrell (2002) criticized the advancement of knowledge in the field regarding multisite evaluation by noting that little had been done to define multisite evaluations, to distinguish among different types of multisite evaluations, or to develop guidelines for determining the most appropriate type of multisite evaluation for a given purpose. Similarly, they note that the process of planning for the challenging context of multisite evaluations was unstructured and completely dependent on the skills of the individual evaluators involved in a particular study.

In response to these continued limitations of understanding, Straw and Herrell offer the framework seen in Figure 4 below as a way for evaluators to begin more systematic planning for variations across different types of multisite evaluations. Within this framework, they define cluster evaluations as evaluations that examine community collaborative projects that report information in a confidential manner to a program funder. Multisite evaluations are defined as varied projects that collect data from local project sites and aggregate data to a national evaluation of sites that have been operational for some time. The third category of their framework, multicenter clinical trials, represents randomized clinical trials conducted across multiple sites for the purpose

of increasing sample size, improving representativeness of the study population, or ensuring that the intervention had effect in multiple contexts.

Dimension	*Cluster Evaluation*	*Multisite Evaluation*	*Multicenter Clinical Trial*
Principal purpose	Examine variation	Estimate impact or examine variation	Estimate impact
Phase of research	Exploratory	Exploratory to confirmatory	Confirmatory
Source of intervention	Program driven	Program, early data driven	Research driven
Standardization	Extensive variation	Some to extensive variation	Extensive standardization
Timing of cross-site evaluation	Planned to retrospective	Planned to retrospective	Planned
Role of cross-site evaluation	Formative	Summative	Summative
Model of cross-site evaluation	Collaborative	National evaluator to collaborative	Collaborative

Figure 4. Straw and Harrell's Framework: Variations across Types of Multi-site Evaluations

Leff and Mulkern (2002) discuss two underlying principles of multisite evaluations. The first principle (science-based) they highlight described the "common call" for science-based practice in program interventions that are publicly funded. In other words, they suggest a means of accountability—where program outcomes are systematically measured and reported upon in order for "successful" programs to be quickly brought to scale for the greatest number of beneficiaries. The second principle (participatory) described in this work implies that stakeholder groups should have meaningful input in all phases of multisite evaluations and highlights the idea that multisite evaluations should reflect the interest and values of various stakeholder groups in public systems, including public officials, evaluators, consumers, advocates, and family members. The idea of stakeholder groups in a multisite setting is different than in

a single program site as there are various stakeholders of the larger funded initiative and stakeholder groups of each local project that makes up the multisite program. Leff and Mulkern (2002) note that many studies that attempted to involve all stakeholders in meaningful ways found the process to be inefficient and time-consuming and that the methods chosen for the multisite evaluation "were driven by the capabilities of the least capable sites" (p. 97). Leff and Mulkern's two principles—science-based practice and the participatory principle—will inform the design of this study, which will look at the nature of involvement in and use of evaluation in multisite settings.

A year later, Lawrenz and Huffman (2003) claimed that little was known about what happens when potential evaluation users at either the project or the larger program level participate in designing and implementing studies or when the results become widely available at the end of project or program funding. Their question about the multitude of potential users is directly related to Leff and Mulkern's participatory principle in that it recognizes the complexity of the stakeholders within a multisite context. Lawrenz and Huffman (2003) go on to assert that in a multisite evaluation setting, the local projects become the "stakeholders" of the larger program while still working to involve local stakeholders in programmatic decision making in the local project. As a result, Lawrenz and Huffman (2003) propose a three-stage process for "multi-site, negotiated centralized" evaluations appropriate to agencies like the National Science Foundation (NSF) that fund large scale programs. The model Lawrenz and Huffman propose calls for a multisite evaluator to: 1) develop local project-level evaluations, 2) negotiate a process for a centralized program evaluation that will meet the

needs of the funding agency, and 3) collaboratively design an integrative evaluation by combining the proposed local and central studies.

Lawrenz and King (2009) continue examination of this topic through a 5-year grant from the National Science Foundation. The "Beyond Evaluation Use" research project is grounded in the NSF context where well funded, large-scale, multi-site evaluation studies are common. The project looked at four NSF programs: Advanced Technological Education (ATE; 1993–2005), Collaboratives for Excellence in Teacher Preparation (CETP; 1993–2000), Local Systemic Change through Teacher Enhancement (LSC; 1995–2005), and Utah State Mathematics and Science Partnerships–Research and Evaluation Teaching Assistance (MSP-RETA; 2002–2006). The ultimate goal of the research is to expand the understanding of evaluation practices that promote evaluation use and influence by contrasting the type and scope of use or influence of different approaches to program evaluation in the multisite context. Their initial hypothesis contends that existing notions of participatory evaluation—where a group of people works together to conduct an evaluation—is not directly applicable to large multisite settings. For example, the multisite program evaluations studied have many layers of stakeholders, which results in sites within the same program being unfamiliar with each other. If the context then affects practices within participatory evaluation, their research proposes that the nature of use or influence would be affected as well.

Toal (2009) and Johnson (2008) were members of the Lawrenz and King team who developed an instrument designed to measure involvement in and use of evaluation in the four multisite NSF programs. Using exploratory factor analysis, Toal (2007)

identified two factors related to involvement (planning and implementation), and Johnson (2008) identified three factors related to use (increased knowledge, understanding, and skills; increased belief in the importance of evaluation; and indications of future use of knowledge gained).

As this literature review has noted about both participatory evaluation and evaluation use, empirical research on multisite evaluations, specifically in science, technology, engineering and math education programs is limited. What does exist tends to rely on case studies and narrative reflections and offers evaluators limited guidance on the critical factors for facilitating effective involvement of stakeholders in a multisite context.

Summary

Using the Cousins and Whitmore framework to examine the research on participatory evaluation, it becomes clear that there is a paucity of empirical research on the topics of interest in this study. This might leave evaluation scholars wondering about many aspects of participatory evaluation like appropriate methods for identifying stakeholders to involve, specific examples of how and when to involve stakeholders, and processes or mechanisms to effectively engage stakeholders in different aspects of evaluation tasks. This dissertation research will address some of these weaknesses in the current literature as it seeks to identify the aspects of participation that result in use in a multisite setting.

Similarly, when scholars review the literature on evaluation use, criticism of the empirical work focuses on a flawed standard of evidence. That is, there is often an over-

reliance on people's self-reported information about use. In addition, these self-reported data are often not validated through other means nor is there systematic triangulation of information.

Finally, this section highlighted that much of the literature on participation and use has come from bounded environments, i.e., single-location local programs. This dissertation research, however, is concerned with the effects of participation in and use of evaluation in large scale, multisite evaluations. Inevitably, this changes the unit of analysis from discrete findings and actions to understanding how an entire network is affected by the infusion of evaluation processes and evaluation findings.

Using the "Beyond Evaluation Use" data, the current study will focus on the participatory frameworks set out by Cousins and Whitmore (1998) and examine the relationship between participatory evaluation and evaluation use in large, multisite evaluations. This study will explore the intersection of Toal and Johnson's findings to examine the relationship of involvement in and use of evaluation in a multisite evaluation context as well as to determine whether the characteristics of involvement and use change and in what ways they change as a result of the multisite context. By using the results from two of the only empirical inquiries related to multisite evaluation (Toal, 2009; Johnson, 2008) and examining the relationship between the two, this research hopes to move knowledge in the field from that of descriptive characteristics to knowledge that facilitates direct activities that promote participation resulting in use.

The next chapter will present a detailed description of the data collection and analysis necessary to untangle this complex relationship. In addition, the chapter will describe the survey and interview participants and data collection procedures.

CHAPTER 3

METHODOLOGY

In this chapter, I discuss the methodology adopted to understand how involvement in evaluation and use of evaluation took place in the cases that were studied. This study is a secondary analysis of mixed method data collected for the Beyond Evaluation Use research project. First, I describe the participants in this study. I then detail the instruments used as primary sources of data. Finally, I describe the data collection and analysis processes.

Participants

The Beyond Evaluation Use research (EUG) team acquired names and email addresses for the entire population of principal investigators (PIs), co-principal investigators (co-PIs), and evaluators for each of the four NSF programs: ATE (1993–2005), CETP (1993–2000), LSC (1995–2005), and MSP-RETA (2002–2006). The list of participants was compiled from project and meeting rosters provided by the program evaluation personnel. Although the rosters contained the most recent emails on record, some names were missing email addresses. In these cases, a member of the EUG team searched the Internet for an email address. The resulting population for this study consisted of all 935 project PIs and evaluators. (Appendix A contains descriptions of the programs and their evaluations). The primary goal of this research was to examine the broad impact of the four program evaluation projects on the field, rather than the uses and influence on NSF as the intended user. While research exists on use and influence of primary intended users through the utilization-focused evaluation principles of Patton

(2008), the current study looked at influence beyond those primary intended users by examining influence on secondary or unintended users.

Each evaluation project operated on a unique timeline for differing durations, with the ATE evaluation project still ongoing in 2009 and the CETP program evaluation running for only 4 years. Additionally, although all projects were about STEM education, each evaluation project examined different aspects of it, ranging from mostly community college-based STEM workforce development, to district based in-service teacher professional development, collaboratives for pre-service teacher education or partnerships encompassing multiple aspects of STEM education. The Beyond Evaluation Use research team prepared detailed case studies of each project.

Primary population. Participants from the four programs share most characteristics of interest but differ on one critical dimension. ATE, CETP, and LSC program evaluations were similarly structured in that there was a distinct program-level evaluation in which projects were asked or required to participate. The Utah State MSP-RETA, however, promoted and provided technical assistance with evaluation for all other MSP projects. Because of this difference, in this study MSP-RETA data were used in limited ways—primarily descriptively. More details regarding this limited use of MSP-RETA data follow later in this section.

ATE, CETP, and LSC were purposefully selected to be representative of the continuum of participation in decision making. The ATE program evaluation provided few opportunities for local participation, while the LSC program evaluation mandated project participation, and the CETP program evaluation offered project involvement

opportunities where individual projects were encouraged, but not required, to participate. By including these three program evaluations, this study encompassed a range of opportunities and experiences found in complex, multi-site evaluations of this nature.

Procedures

Survey Process. An online survey was conducted with evaluators and project leaders in program-level evaluations of the four large, multi-site science, technology, mathematics and engineering education interventions. The EUG team administered the survey with a staggered release via email by program to minimize technical difficulties associated with mass emails (e.g., SPAM filter delivery interference, returned emails, server overload). While distribution was staggered, all programs received an invitation email that contained a description of the survey, an explanation of its use, and the Human Subjects approval number and contact information to call with questions. The email also contained a link to the online survey with a unique login and password. Two and four weeks following the first email, the team sent a reminder to non-respondents. Initial recipients of the first invitation brought questions that prompted the EUG team to make clarifications to the invitation regarding the evaluation of most interest (program-level, not project level). In addition, the team shortened the reminder emails for the remaining three projects (CETP, ATE, MSP-RETA; see Appendix B for the survey email invitations and reminders). The data were collected from August 10, 2006 through January 17, 2007.

Twenty-two percent (215/935) of the emails were returned as undeliverable. EUG Team members conducted Internet searches and, in some cases, phone research to find working email addresses. About half (109/215) of the invalid email addresses were

replaced with new emails. Of the new emails, 86 worked and 23 failed again. The final sampling frame, the number of ATE, CETP, LSC, and MSP-RETA PIs and evaluators with valid email addresses, was 810, with an overall response rate of 46% (372/810). Three respondents provided unusable data that were not used in the final analysis (see Table 4).

Table 4.

Survey Respondent by Role

Survey Respondent Role	ATE	CETP	LSC	MSP-RETA	Total Respondents
Evaluators	11	19	30	3	63
Non-Evaluators (PIs, project staff, project-related consultants, etc.)	175	35	43	53	306
Total	186	54	73	56	369*

* Two respondents from ATE and one from LSC provided unusable data, so their surveys are not included in the analysis.

Respondents came from all but three states in the U.S (Arkansas, Kansas, and Wyoming), including the District of Columbia and Puerto Rico. Fifty-four percent of the population was male, 43% female, and 3% preferred not to respond to the gender question. Data from three respondents had to be discarded due to technical problems or a clear misunderstanding of the survey's focus as indicated in the open-ended comment sections.

The EUG team made efforts to reduce error in the survey design. Dillman (2007) identifies four types of error common in survey research:

1. Sampling error, which results from surveying only a subset of the population;
2. Coverage error, which results from not everyone having a random chance of being included in the sample;
3. Measurement error, which results when questions are misunderstood by the survey-taker; and,
4. Non-response error, which occurs if individuals who did not respond are different than those who responded in ways that are relevant to the study.

Sampling error is minimized in this study because the team surveyed the entire population, except for those with invalid email addresses. In other words, the EUG team did not use any sampling techniques to choose a subset of the population to survey; instead, the EUG team surveyed as many of the population as possible. In addition, the EUG team reduced measurement error by careful survey construction and the use of think alouds, described later in this chapter.

This survey is likely impacted by some coverage error because of the inability to acquire valid email addresses for the entire population. Still, the team accepted this as a consequence of the decision to distribute the survey via email. The EUG team reasoned that postal addresses would likely be more costly to use and result in similar error because address information may also be outdated or more difficult to obtain. In addition, the population was all affiliated with NSF projects and regular use of email was likely because NSF funding is awarded via an online application process that requires email addresses of applicants.

With a response rate of 46%, there was the possibility for non-response error. The EUG team compared survey respondents and non-respondents to determine any significant differences between the two groups that might have been related to their levels of involvement. A random sample of non-respondents was contacted by email first, then by telephone if email contact did not solicit a response. The simplified survey asked for information related to three items on the original instrument: 1) overall involvement in the program evaluation, 2) overall impact of the process, instruments, or results, and 3) the reason for not responding to the initial survey. Since the original instrument did not ask an overall level of involvement question, the mean of all 13 involvement items provided the comparison measure. Results indicated that non-respondents did not have significantly different levels of involvement.

The items in the larger survey related to use were distilled into a single question and resulted in the choice of the word "impact" as a global concept to encompass questions about use of the process, instruments, and results of the evaluation. The analysis of the impact question showed a significant difference between respondents and non-respondents in two of the four programs. Table 5 below shows the respondents' and non-respondents' means for overall involvement and use/impact. As a result, the difference between respondents and non-respondents on this combined item may indicate the possibility of upward bias in the estimates of overall use for survey respondents, or it may highlight the possibility of overlap in interpretation of the terms use, influence, and overall impact.

Table 5.

Summary of Findings of Non-Response Analysis

	Involvement Mean			Use/Impact Mean		
NSF Program	Respondents	Non-Respondents	p-value	Respondents	Non-Respondents	p-value
ATE	2.39	2.36	.913	2.50	2.06	.03
CETP	2.38	1.82	.066	2.52	1.76	.01
LSC	2.71	3.14	.147	2.85	2.69	.70
MSP	1.76	1.95	.493	2.02	1.90	.48

Project Interviews. Data were also collected from survey respondents using in-depth interviews. The interviews were conducted with two purposefully selected samples of project leaders and evaluators who had responded to the survey. One sample contained respondents with varying reported levels of involvement and use, e.g., low use—high involvement, while the other sample was entirely high users. To purposefully identify the sample, the EUG team members identified potential interviewees based on overall involvement and use mean scores to include respondents with a range of involvement and use (see Table 6).

Table 6.

Description of Interviewees

Interviewees	Role	Program	Level of Involvement (Mean of 13 related items)	Level of Use (Mean of 13 related items)
1	PI	ATE	1.00	3.00
2	Evaluator	ATE	3.69	4.00
3	Evaluator	ATE	3.92	4.00
4	PI	ATE	2.00	1.15
5	PI	ATE	3.00	1.44
18	PI	ATE	3.08	3.31
6	PI	CETP	1.54	1.37
7	Evaluator	CETP	3.46	3.39
8	Evaluator	CETP	4.00	2.65
9	PI	CETP	3.77	1.91
10	PI	CETP	1.85	3.00
19	PI	CETP	1.15	3.14
20	PI	CETP	3.31	2.94
11	PI	LSC	3.46	3.84
12	PI	LSC	2.38	3.94
13	PI	LSC	1.15	2.04
14	Evaluator	LSC	3.69	1.26
15	Evaluator	LSC	3.69	3.74
16	Evaluator	LSC	2.69	3.65
17	PI	LSC	2.92	3.59

Twenty interviews were conducted with six to seven interviews per evaluation project. Interviews lasted between 30–50 minutes. All were taped and transcribed. The interviews covered involvement and use as well as background information, motivation for involvement, perceived influence on the evaluation, and the ways in which the evaluation was used or had influence. (Please see Appendix C for the interview consent form and Appendix D for the interview protocol).

Scale Development

The EUG team began survey and scale development by brainstorming over 200 items related to use, influence, and involvement. The challenge was to develop an instrument that was of reasonable length and still adequately able to cover the three constructs. The team worked to narrow down choices and did pilot testing until they were satisfied with the wording of the items.

Involvement. Toal (2007), a member of the EUG research team, led efforts to develop and validate the involvement scale. Based on work of Lawrenz and Huffman (2003), she suggested that both Cousins and Whitmore's (1998) three-dimensional conceptualization of participatory evaluation and Burke's (1998) key decision points may help evaluators better understand the participatory nature of multi-site evaluations. Therefore, the EUG team incorporated these frameworks into items that aligned with notions of involvement and use.

The first two dimensions on Cousins and Whitmore's (1998) three-dimensional conceptualization (control of the evaluation and stakeholder selection) were previously established by the multi-site evaluators with the design of the program evaluations.

Specifically, the program evaluators pre-determined the control of the evaluation. Stakeholder selection also did not require an item on the scale because all of the evaluations were multi-site and stakeholders were limited to the individual projects within each program. Therefore, stakeholder selection was held constant for each evaluation. As a result, the Evaluation Involvement Scale only measures Cousins and Whitmore's (1998) third dimension: depth of participation.

Using Burke's key decision points as a general framework, the scale measured depth of participation by the number and extent to which stakeholders took part in the key decisions. The EUG team aligned each of Burke's key decisions to one of three evaluation stages: evaluation planning, implementation, and communication of results. Based on the EUG team's evaluation expertise, the team added several items related to each of the three evaluation stages in order either to make finer distinctions between the decision points or to make the questions more clear. Each involvement item offered the following response options:

- No
- Yes, a little
- Yes, some
- Yes, extensively
- I don't think this activity took place

In the analysis, "No" = 1, "Yes, a little" = 2, "Yes, some" = 3, and "Yes, extensively" = 4.

Items were developed to align with Burke's decision points although the team made modifications in two areas. First, the team further divided the decision points Burke called "collecting the data" and "synthesizing, analyzing, and verifying the data" decision points into six specific items on the survey. In addition, Burke labeled the final evaluation

phase as "communication of evaluation findings," but the EUG team used Burke's (1998) "developing action plans for the future," but added "based on the results" to be more specific. The team also replaced Burke's "controlling and using outcomes and reports" with three tasks related to decision making (see Table 7 for the involvement items).

Table 7.

Involvement Items

The items related to evaluation planning were:
• I was involved in the discussions that focused the evaluation. • I was involved in identifying evaluation planning team members. • I was involved in developing the evaluation plan.
Items related to collecting data and synthesizing, analyzing and verifying data were:
• I was involved in developing data collection instruments. • I was involved in developing data collection processes. • I was involved in collecting data. • I was involved in reviewing collected data for accuracy and/or completeness. • I was involved in analyzing data. • I was involved in interpreting collected data.
Items related to the communication of evaluation findings were:
• I was involved in writing evaluation reports. • I was involved in reviewing evaluation reports for accuracy and/or completeness. • I was involved in presenting evaluation findings (e.g., to staff, to stakeholders, to an external audience). • I was involved in developing future project plans based on evaluation results.

Toal's analysis (2007, pg. 65) found that the items were more than adequate to measure the construct of involvement when examined using Cronbach's coefficient alpha (α= 0.943).

Evaluation use. No validated measures of use exist, so the EUG team created survey items to measure use and influence. The development of items regarding the use construct was not guided by a framework, but rather on a thorough review of the literature regarding evaluation use, applied evaluation experience of team members, and the input of respected evaluation experts (see Table 8 for the survey items related to the evaluation use construct).

Table 8.

Evaluation Use Survey Items

Evaluation Use Survey Items
• The evaluation increased my knowledge/understanding of how to plan an evaluation (e.g., discussing the focus of the evaluation, identifying evaluation planning team members, developing the evaluation plan).
• The evaluation increased my knowledge/understanding of how to implement an evaluation (e.g., developing data collection instruments and processes, collecting, analyzing, reviewing, and interpreting data).
• The evaluation increased my knowledge/understanding of how to communicate evaluation findings (e.g., developing future plans for your project, writing and reviewing evaluation reports, and presenting evaluation findings).
• The evaluation increased my knowledge/understanding of Science, Technology, Engineering, and Mathematics (STEM) education.
• The evaluation increased my knowledge/understanding of STEM education evaluation.
• The evaluation increased my knowledge/understanding of my project.
• The evaluation improved my skills in planning an evaluation (e.g. discussing the focus of the evaluation, identifying evaluation planning team members, developing the evaluation plan).
• The evaluation improved my skills in implementing an evaluation (e.g. developing data collection instruments and processes, collecting, analyzing, reviewing, and interpreting data).
• The evaluation improved my skills in communicating evaluation findings (e.g., writing and reviewing evaluation reports, presenting evaluation findings).
• The evaluation improved my skills for working on my project.
• The evaluation increased my belief in the importance of planning an evaluation (e.g., discussing the focus of the evaluation, identifying evaluation planning team members, developing the evaluation plan).
• The evaluation increased my belief in the importance of implementing an evaluation (e.g., developing data collection instruments and processes, collecting, analyzing, reviewing, and interpreting data).
• The evaluation increased my belief in the importance of communicating evaluation findings (e.g., developing future plans for your project, writing and reviewing evaluation reports, presenting evaluation findings).
• The evaluation increased my belief in the importance of STEM education.
• The evaluation increased my belief in the importance of STEM education evaluation.
• The evaluation increased my belief in the importance of my project.

The response options offered for the use items matched the response options for the involvement items (No; Yes, a little; Yes, some; Yes, extensively) and used the same numeric coding for analysis. Johnson (2008) led efforts to construct and validate the use/influence scale. Johnson's analysis found that the items were more than adequate to measure the construct of use/influence when examined using Cronbach's coefficient alpha (α=.971). More details of the validation process follow in the Exploratory Factor Analysis section below.

Think Alouds. Prior to finalizing the survey, the EUG team conducted three think alouds, otherwise known as cognitive interviews (Dillman, 2007). Three doctoral students in Evaluation Studies, all having experience in the field of evaluation, took the survey with one of the EUG team members. The team member asked the participants to verbalize their thought process as they responded to the survey. The team member took notes on each participant's reactions to the survey's content, format, and ease of understanding. After the participant completed the survey, the EUG team member discussed how the survey items could be improved. After each interview, the EUG team member shared suggestions with the group and made changes accordingly. As part of the validation process for the survey overall, each of the participants answered general questions about the survey.

Expert opinion. Over the course of 1 year, the EUG team extensively reviewed and deliberated over which activities adequately covered the constructs of evaluation involvement and use. The EUG team included two internal experts: University of Minnesota faculty members, Drs. Jean A. King and Frances Lawrenz, both of whom are

well versed in the evaluation literature. Dr. Lawrenz has also been an NSF evaluator for over 35 years and was personally involved in each of the program evaluations included in this research. Her involvement in the evaluations helped ensure that items were relevant and inclusive. The EUG team began compiling potential items in mid-summer, 2005 and continued revising items until July 2006.

In May 2006, the survey draft, almost complete at that time, was sent via email to five external experts: the program evaluators for ATE, LSC, and MSP-RETA (Arlen Gullickson, Iris Weiss, and Catherine Callow-Heusser) and Michael Quinn Patton, an advisory board member. Each expert sent detailed reactions and minor suggestions back to the EUG team by June 2006. The EUG team considered all of their suggestions and implemented those they deemed appropriate. The final version was sent to Marvin Alkin, the EUG team's meta-evaluator, also an external expert, in July 2006. He proposed minor revisions related to format and word choice that were implemented, and the survey wording was finalized by August 2007.

Exploratory Factor Analysis. Two EUG team members conducted initial cleaning and analysis of the survey data. In addition, Toal (2007) conducted exploratory factor analysis of items related to involvement and Johnson (2008) conducted exploratory factor analysis of items related to use and influence. Toal submitted the thirteen items related to involvement to EFA using principal components analysis with Varimax rotation. Eigenvalues greater than 1.0 supported the decision to retain two factors that explained 72% of the variance. Items that loaded at least 0.4 on a factor and were more than 0.2 higher than another factor loading were retained. This rule supported the removal of the

following two items: reviewing evaluation reports for accuracy and/or completeness, and developing future project plans based on evaluation results. The retained factors related to involvement were:

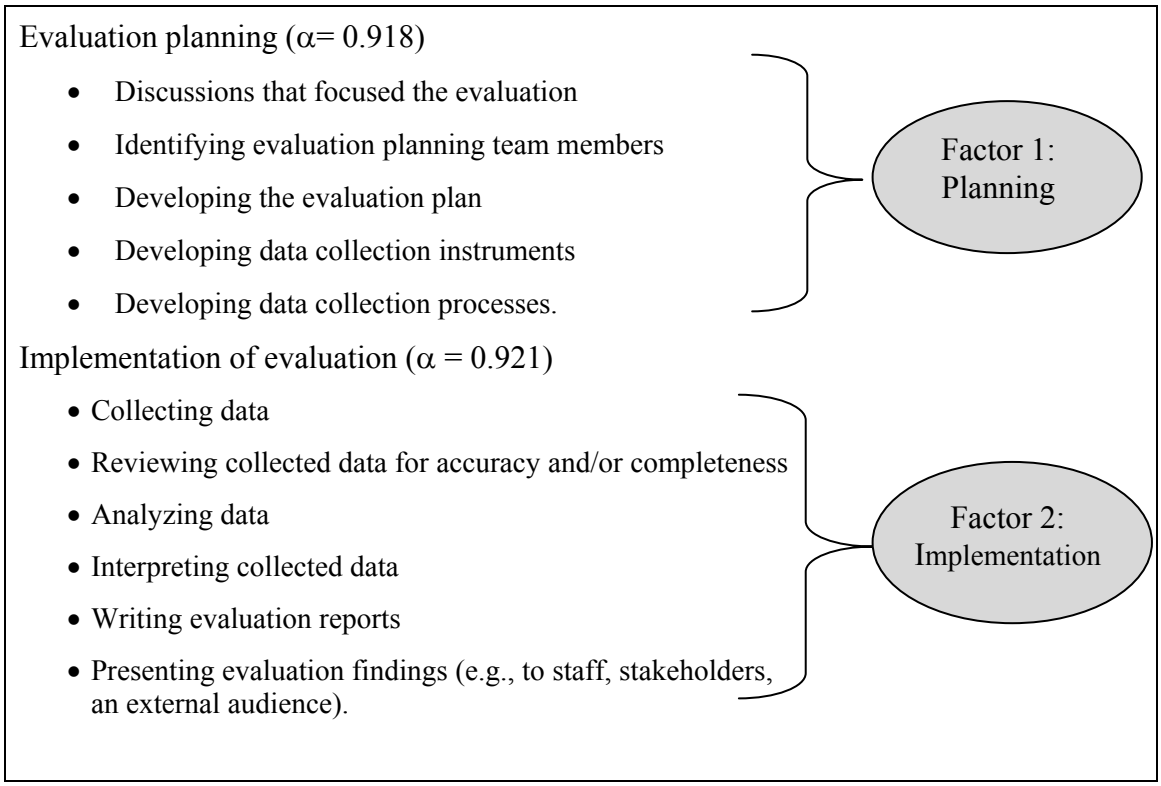


Figure 5. Toal's involvement items aligned with retained factors

In addition, Johnson submitted the 23 items related to use/influence to EFA using principal components analysis with Varimax rotation. Eigenvalues greater than 1.0 supported the decision to retain three factors that explained 78% of the variance. Items that loaded at least 0.4 on a factor and were more than 0.2 higher than another factor loading were retained. This rule supported the removal of the following two items:

improved my skills as a STEM educator, and improved my skills as a STEM education evaluator. The retained factors related to use/influence included:

- Increased knowledge, understanding, and skills
- Increased belief in the importance of evaluation
- Indications of future use of knowledge gained

Figure 6 shows how the original items aligned within the three factors related to use/influence.

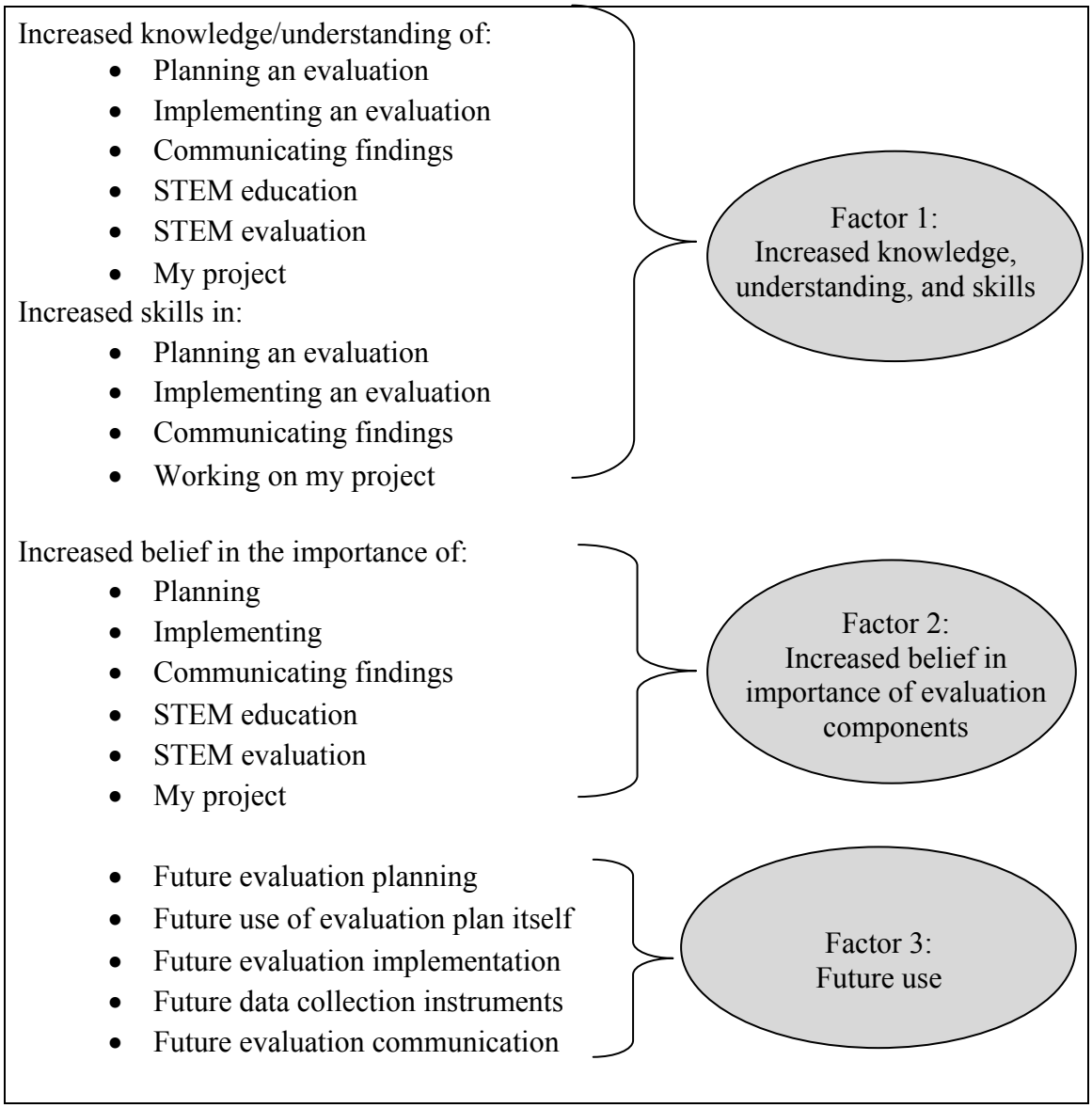


Figure 6. Johnson's use/influence items aligned with retained factors

Rationale for Research Design

A research design is a logical structure that links data collection, data analysis, and its findings to the purpose of the study (Yin, 2003). As noted in Chapter 1, this study explored the nature of the relationship between participatory evaluation and evaluation

use in a multisite evaluation context. Based on the purpose of this study, there is one primary and three supporting research questions:

Primary Question 1: What is the nature of the relationship between participation in and use of evaluation in a multisite evaluation setting?

Supporting Question 1: To what extent can evaluation use be explained by participation in multi-site evaluations? And to what extent can participation in evaluation be influenced by expectations of use in multi-site evaluations?

Supporting Question 2: What contribution does any single empirically derived factor related to participation in evaluation make to the explanatory power of participation? What contribution does any single empirically derived factor related to use make to the explanatory power of use?

Supporting Question 3: Do certain empirically derived factors related to participation in evaluation contribute to explaining the composite of factors that make up the idea of use? Do certain empirically derived factors related to use of evaluation contribute to explaining the composite of factors that make up the idea of participation?

Many scholars purport that mixed methods yield unique insight from qualitative or quantitative studies conducted independently (Chelimsky, 1997; Patton, 1997; Greene, 2007). Given that this study seeks to both explore and explain involvement in and use of evaluation in multisite evaluations, a mixed method approach best aligned with the desired outcomes of this study. The study used a sequential component design that implemented quantitative analysis first, then qualitative analysis. As such, qualitative data are used to shed light on quantitative findings at the interpretation phase of this study

and serve to confirm or challenge quantitative findings. The data analysis process will be discussed in detail in the following section.

Data Analysis Procedures

This study initially conducted canonical correlation analysis of the involvement and use scales for all three programs combined. Canonical correlation analysis (CCA) studies associations between two sets of random variables (Thompson, 1984). More specifically, CCA is a multivariate statistical model that facilitates the study of interrelationships among multiple dependent variables and multiple independent variables. Multivariate modeling acknowledges that in real-world applications, there are multiple causes and effects and as such, using multivariate statistical methods allows a researcher to shed light on how each variable contributes to the relation of all variables, particularly in early descriptive examination of a phenomenon. In addition, multivariate analysis does not increase the risk of Type I error rates that would likely occur if multiple univariate tests were used to examine the same data. CCA identifies components of one set of variables that are most highly related linearly to the components of the other set of variables. Figure 7 illustrates the constructs of involvement and use for this study.

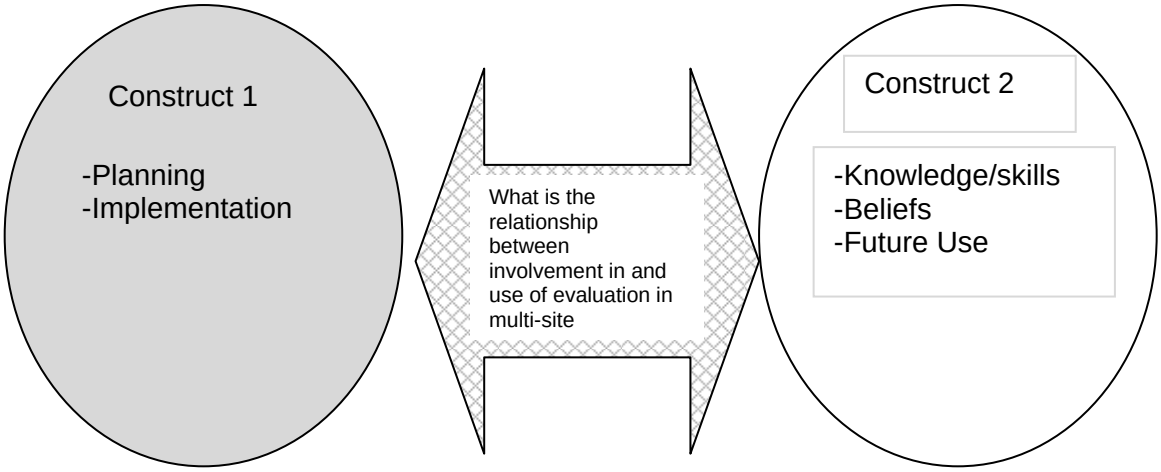


Figure 7. Variable sets related to involvement and use

CCA follows a similar procedure as factor analysis but tries to account for the maximum amount of correlation between the two sets of variables rather than within pairs of single variables. As such, the first canonical function is derived so as to have the highest inter-correlations possible between the two sets of variables. The second canonical function will exhibit the maximum amount of relationship between the two sets that was not accounted for by the first function. In this study then, the functions maximize the correlation between use and involvement by weighting the factors that make up the constructs of use and involvement. The maximum number of canonical functions is equal to the number of variables in the smaller variable set. This study will thus derive two canonical functions since there are two constructs.

Canonical correlation analysis delivers a complex set of results. This research used several criteria to determine significance. First, the overall level of significance of each canonical correlation was examined, using .05 as the level that was minimally

acceptable for interpretation. The remaining canonical coefficients produced through this method of analysis can be cumbersome to interpret. This could result in misinterpretation unless care is given to align results with the intended research questions (Thompson, 1984). This research planned a priori to use several coefficients for interpretation that best aligned with the research questions. Table 9 shows how the coefficients relate to the proposed research questions.

Table 9.

Canonical Coefficients Used For Interpretation

Primary Question 1: What is the nature of the relationship between participation in and use of evaluation in a multisite evaluation setting?			
In other words…	Coefficient used	Research question addressed by coefficient	What does this statistic tell us?
Is there a relationship between involvement in and use of evaluation?	Squared canonical coefficient	Supporting Question 1: To what extent can evaluation use be explained by involvement in multi-site evaluations?	Shows shared variance of use and involvement or the degree to which the two constructs are related
	Redundancy coefficient		A summary measure of the ability of the involvement factors (taken as a set) to explain variation in the use factors (taken one at a time).
And if there is a relationship, where does it come from?	Squared structure coefficient	Supporting Question 2: What contribution does any single factor related to participation in evaluation make to the explanatory power of participation? What contribution does any single factor related to use make to the explanatory power of use?	The contribution any one involvement factor makes in explaining involvement. The contribution any one use factor makes in explaining use. OR the relationship of the factors within a canonical set
	Squared index coefficient	Supporting Question 3: Do certain factors related to participation in evaluation contribute to explaining the composite of factors that make up the idea of use? Do certain factors related to use contribute to explaining the composite factors that make up the idea of involvement?	The contribution any one factor related to involvement makes in explaining use. The contribution any one factor related to use makes in explaining involvement. OR the relationship of each factor in one canonical set to the construct represented by the other canonical set.

The squared canonical coefficient (Rc^2) represents the amount of variance in one set of variables accounted for by the other set of variables and can be used as an indication of relative importance of the independent variate in determining the value of the dependent variate. The value of this coefficient can only range between 0 and the number of variables in the smallest set (in this case 2). This will allow us to judge both the direction of the relationship and the strength of the relationship. Because the Rc^2 coefficient provides an estimate of the shared variance between the sets of variables, it best allows this study to answer the question that focuses on how involvement in evaluation may explain utilization (or how use of evaluation influences involvement). To aid in further interpretation, it is also useful to examine the redundancy (Rd) coefficient. The Rd coefficient provides a summary measure of the ability of a set of independent variables (taken as a set) to explain variation in the dependent variables (taken one at a time).

The squared structure coefficient can help explain the relative contribution of each variable to the variable set of which it is a part. This statistic will allow this study to shed light on the contribution a single factor related to participation in evaluation makes to the explanatory power of use.

The squared index coefficient indicates how closely the factors related to participation resemble a straight line or how closely the factors related to use resemble a straight line. The coefficient ranges from 0 meaning no apparent correlation to the straight line, up to 1 for perfect correlation. This information provides a general measure

of how each factor in one canonical set related to the construct represented by the other canonical set in its entirety.

When analysis of the aggregate data set was concluded, the same procedures will be used to split the data file by program and conduct the analysis on each program. Because the programs each fall at different point along the Cousins and Whitmore continuum regarding the control of the evaluation, this second phase of analysis has the potential to yield helpful insights in understanding the relationship between involvement and use.

Finally, after completing quantitative analysis, this study examines interview data as a means for explaining and interpreting the findings of the CCA analysis. Using Nvivo 8, I will identify descriptive themes from the interviews that support or contradict the quantitative findings and interpret the connection between the interview themes and the CCA results.

IRB Approval

This study was exempt from full IRB review under federal guideline 45 CFR Part 46.101 (b) category 4 (use of existing data). Data for the present study (#1001E75715) were collected for the Beyond Evaluation Use research project at the University of Minnesota.

Chapter 4 will provide a detailed explanation of the results of these analyses.

CHAPTER 4

FINDINGS

This chapter presents the results of an analysis of quantitative and qualitative data used for this study. Chapter 5 will provide an explanation of how the results are relevant to the research questions and discuss the extent to which the results describe or explain the relationship between involvement in and use of evaluation in a multisite evaluation context.

This study sought to examine the relationship between two theoretical multi-dimensional constructs: involvement in, and use/influence of evaluation. This can only be accomplished quantitatively if the constructs are measured. In research led by Lawrenz and King (2009) and completed by Toal (2008) and Johnson (2008), the two constructs were measured using data collected with a survey. It became evident with the survey results that the constructs were multidimensional in nature; Toal identified two factors related to involvement in evaluation, and Johnson identified three factors related to use/influence of evaluation. This multidimensionality made simple correlations between the constructs inappropriate. Because of this, canonical correlation analysis (CCA) was performed to assess the nature of the relationship between the theoretical constructs. CCA allowed the examination of the relationship between the set of dimensions pertaining to each construct simultaneously. The analysis treated the sets of factors identified by Toal and Johnson as dimensions of the constructs and used the factor scores that were computed for the factors as measures of each dimension. In addition to looking at the dimensions of each construct simultaneously, CCA also computes all the bivariate

correlations between individual factors that make it possible to better understand the overall relationship.

According to Thompson (1984), canonical correlation analysis requires three assumptions be met. First, the measurement error of the variables is minimal. Second, variances of the variables are not restricted and third, the magnitudes of the coefficients in the correlation matrix must not be attenuated by large differences in the shapes of the distributions for the variables. These assumptions of linearity, multivariate normality, and homoscedasticity were checked by evaluating the bivariate scatterplots of the canonical function scores using PASW 18.0 (Appendix E). Upon examination of the scatterplots, these data appear to satisfy the assumptions necessary to continue with CCA. While the remaining analysis of survey data used for the study was conducted using SAS 9.0, PASW 18 was used strictly to generate scatterplots to check assumptions because SAS does not offer this capability.

Canonical correlations are the correlations between a linear combination of the involvement factors and a linear combination of the use/influence factors. As such, canonical correlation analysis determines vectors (i.e., the linear combinations) that represent the best possible combination of involvement and use/influence by plotting the shared variance of combinations of the factors. These vectors are often referred to as canonical functions. Canonical functions help us determine the strength of the relationship between two sets of variables (Hair et al., 1988).

CCA Results

Canonical correlation analysis delivers a complex set of results in addition to the main result, which is the correlation between involvement and use/influence. This section will discuss the results as evidenced by a number of indicators and make connections between these indicators and the research questions. The section starts out by discussing the results of the bivariate correlations within and between factor sets, moves on to talk about the standardized canonical coefficient that is the main result, and concludes by discussing the ways in which the squared structure and the squared index coefficients help bring meaning to the standardized canonical coefficient.

The first step in the canonical analysis was to examine all possible bivariate correlations. In other words, correlations were calculated for all paired combinations of the two involvement factors and the three use/influence factors. The results are highlighted in Table 10. Using Cohen's guidelines for interpretation (small = .10 to .30; medium = .30 to .50; large = .50 to 1), we see that there is a medium correlation between the two involvement factors: planning and implementation. The three correlations of pairs within the factors related to use all show large correlations. In addition, the six correlations of pairs across the involvement factors and the use/influence factors all show medium correlations. This provides reasonable confidence that the factors identified by Toal and Johnson adequately reflect the constructs of involvement and use for the purposes of this study because having medium to large correlations of pairs within the constructs is consistent with the factors being related but distinct as would be expected from the factor analyses. The pairs of correlations within the construct of involvement

and within the construct of use/influence are generally higher than the pairs of correlations between (or across) the constructs, which, in turn, supports the idea that involvement and use/influence are separate constructs.

Table 10.

Correlations Within and Across Factor Sets Using All 5 Factors

Correlations Within and Across Factor Sets Using All 5 Factors	Correlation
Correlation of variables within their variable sets	
Planning—Implementation	.4702
Knowledge–Beliefs–Future Use	
Knowledge–Belief	.6837
Knowledge–Future Use	.7078
Belief–Future Use	.5429
Correlations of variables across variable sets	
Knowledge–Planning Correlation	.4053
Belief–Planning Correlation	.3355
Implementation–Knowledge Correlation	.5157
Implementation–Belief Correlation	.3086
Planning–Future use correlation	.3089
Implementation–Future use correlation	.4359

Next, the canonical analysis determines a number of statistics that serve as significance tests. The analysis yielded two functions. First, the canonical correlations of each of the two canonical functions are determined, followed by multivariate tests of both

functions simultaneously. The test statistic employed is Wilks' lambda, which indicates whether the canonical functions, taken collectively, are statistically significant. The final step was to perform redundancy analyses on both canonical functions. Table 11 is an excerpt of an earlier table that contains a summary of what the squared canonical coefficient and redundancy coefficient tell us.

Table 11.

Summary of Squared Canonical Coefficient and Redundancy Coefficient

Guiding question	Coefficient used	Research question addressed by coefficient	What does this statistic tell us?
Is there a relationship between involvement in and use of evaluation?	Squared canonical coefficient	Supporting Question 1: To what extent can evaluation use be explained by involvement in multi-site evaluations?	Shows shared variance of use and involvement or the degree to which the two constructs are related
	Redundancy coefficient		A summary measure of the ability of the involvement factors (taken as a set) to explain variation in the use factors (taken one at a time).

In this study, the standardized canonical coefficient for Function 1 was .556 (31% overlapping variance); Function 2 was .15 (2% overlapping variance). Chi-square tests of the canonical coefficients were done to establish whether or not the paired observations were independent of each other and indicate whether other correlations from that point on were significant, even after removing variance accounted for by the correlations for

Function 1 and 2. With both canonical functions included, λ^2 (6) = 49.02, p<.001, and with the first function removed λ^2 (2) = 2.87, p = .238. This means that the canonical correlations for the two functions together were statistically significant but when the first function was removed the second function was not statistically significant. Upon further examination of the data, it was found that over 180 of the cases were dropped from the analysis because of missing data. This occurred because the items used to determine the third use/influence factor (future use) were not answered by a large subset of respondents, as the items did not apply to them. Given this, it seemed important to repeat the analysis, removing the future use factor from the use/influence variable set in order to maximize the power of the 313 responses.

As a result, the data including both involvement factors and *two* of the use/influence factors were analyzed. As shown before for the original analyses, the canonical analysis first produced all paired correlations within and across the constructs of involvement and use/influence. The results are presented in Table 12. As can be seen, the pattern of stronger correlations of factors within constructs and lower correlations across constructs exhibited in the original analysis was also true in this second analysis.

Table 12.

Correlations Within and Across Variable Sets Using Four of the Factors

Correlations Within and Across Variable Sets Using Four of the Factors	Correlation
Correlation of variables within their variable sets	
Planning–Implementation	.6056
Knowledge–Belief	.7696
Correlations of variables across variable sets	
Knowledge–Planning Correlation	.4590
Belief–Planning Correlation	.3585
Implementation–Knowledge Correlation	.5221
Implementation–Belief Correlation	.3249

Once again, two canonical functions were created representing the strength of the relationship between the two constructs. The standardized canonical correlation coefficient for the first function was .558 (31% overlapping variance); the second was .13 (2% overlapping variance). These were very similar to the results from the original analysis; however, the results of the chi square testing were different. With both canonical functions included, $\lambda^2 (4) = 103.22$, $p<.001$, and with the first function removed $\lambda^2 (1) = 4.79$, $p = .029$. By removing the factor related to future use from the model, the canonical correlation values were not greatly affected but now both canonical correlations were statistically significant based on the Chi-square test, meaning that both functions show a relationship between involvement and use. However, in order to get a

better sense of whether the results have practical significance, it is important to examine several additional statistics delivered by CCA.

The standardized canonical coefficients for each of the factors within the two canonical functions are shown in Table 13. From these coefficients we can determine the relative influence of the factors in the creation of the canonical functions. The strength of the coefficient helps us determine which factors the most in the creation of the function. The standardized coefficients in this study may range from 0 to 2, because of the number of variables in the set with the fewest variables, and represent the weights that would be assigned to the factors if the factor scores were normalized. Given that, Function1seems to consist primarily of the knowledge, understanding, or skill factor and the implementation factor. This determination was made by looking at the value of the positive coefficients under Function1for each variable set and determining which factor(s) for each variable set has the larger coefficient. As such, Function 2 seems to represent a more balanced contribution by all four factors. Regarding the two negative coefficients under Function 2, Frederick (1999) recommends that any negative function coefficients be interpreted as zero. This assertion is also supported by Sherry and Henson (2005) who claim that the only time a negative coefficient is interpreted in CCA is when examining the structure coefficients (explained next).

Table 13.

Standardized Canonical Coefficients

Factor Name	Canonical Coefficients Function 1	Canonical Coefficients Function 2
Involvement factors		
Planning	.36	1.20
Implementation	.74	-1.02
Use/Influence Factors		
Knowledge, understanding or skills	1.17	-1.04
Belief in the value of evaluation	-.24	1.54

While the canonical coefficients allow us to begin describing the strength of the relationship between the two constructs, we can learn more intricate details about the relationship by looking at two other important statistics delivered by CCA. The squared structure coefficient allows us to speak in more detail about the nature of the factors within the construct to which they belong, and the squared index coefficients allow us to look across the constructs. Table 14 is an excerpt of an earlier table that reminds us what this coefficient helps us interpret about the relationship between use and involvement.

Table 14.

Summary of Squared Structure Coefficient Use

Guiding question	Coefficient used	Research question addressed by coefficient	What does this statistic tell us?
And if there is a relationship, where does it come from?	Squared structure coefficient	Supporting Question 2: What contribution does any single factor related to participation in evaluation make to the explanatory power of participation? What contribution does any single factor related to use make to the explanatory power of use?	The contribution any one involvement factor makes in explaining involvement. The contribution any one use factor makes in explaining use. OR the relationship of the factors within a canonical set

The squared structure coefficients examine the relationship of each single factor within the set of which it is a part. Details regarding the structure coefficients are found in Table 15. In Function 1, we see that implementation contributes the most to the definition of the involvement construct though planning is still strong; planning does contribute more substantially than implementation for involvement in Function 2. Regarding the construct of use/influence, we see that Function1highlights the sizable contribution of the knowledge, understanding, or skills factor; whereas Function 2 shows the sizable role played by belief in evaluation in defining the construct.

Table 15.

Squared Structure Coefficients*

Factor Name	Structure Coefficients Function 1	Structure Coefficients Function 2
Involvement factors		
Planning	.81	.57
Implementation	.96	-0.29
Use/Influence Factors		
Knowledge, understanding or skills	.99	.15
Belief in the value of evaluation	.66	.75

* These coefficients help us see relative contribution of factors to their own set.

Finally, the squared index coefficients allow us to understand more about the extent a single factor contributes to explaining the composite of the factors that make up the other construct. We see a summary of how this coefficient is used in interpretation in Table 16.

Table 16.

Summary of Squared Index Coefficient Use

Guiding question	Coefficient used	Research question addressed by coefficient	What does this statistic tell us?
And if there is a relationship, where does the relationship come from?	Squared index coefficient	Supporting Question 3: Do certain factors related to participation in evaluation contribute to explaining the composite of factors that make up the idea of use? Do certain factors related to use contribute to explaining the composite factors that make up the idea of involvement?	The contribution any one factor related to involvement makes in explaining use. The contribution any one factor related to use makes in explaining involvement. OR the relationship of each factor in one canonical set to the construct represented by the other canonical set.

The index coefficients in Table 17 appear to be more conservative than the structure coefficients seen above. For this reason, Thompson (1984) sometimes finds them to be a more solid base for interpretation. From the squared index coefficients below, we see that in Function 1, implementation contributes slightly more than planning to explain the relationship to use/influence; likewise the knowledge, understanding, or skills factor contributes the most to explaining use's relationship with involvement. The first function also shows a stronger relationship (based on the values of the coefficients) than the second function. This would be expected because of the way in which CCA places the best possible linear combination of variables in Function1and aligns remaining variables in such a way as to maximize the remaining correlation in the remaining functions. In Function 2, we see that planning contributes the most to explain

involvement's relationship with use; whereas belief in the role of evaluation contributes the most to explaining the relationship with involvement.

Table 17.

Squared Index Coefficients*

Factor Name	Index Coefficients Function 1	Index Coefficients Function 2
Involvement factors		
Planning	.45	.08
Implementation	.53	-0.04
Use/Influence Factors		
Knowledge, understanding or skills	.55	.02
Belief in the value of evaluation	.37	.10

* These coefficients help us see the relationship of individual factors with the other complete set of factors.

Because the three programs examined in this study represented different points along the Cousins and Whitmore's (1998) depth of participation dimension the final step in CCA was to split the file by program and examine the results for any differences by program. The depth of participation was characterized by Cousins and Whitmore as a continuum ranging on one end from consultation with no decision-making control or responsibility to deep participation that involves stakeholders in all aspects of the evaluation including design, data collection, analysis, reporting and dissemination. As such, the ATE program provided fewer opportunities for local projects to be involved in evaluation planning, the LSC program mandated project participation but afforded only

limited decision-making about the evaluation to the local projects, and the CETP program offered project involvement opportunities for local projects to collaboratively plan aspects of the multi-site evaluation.

Once again two canonical functions were created representing the strength of the relationship between the two constructs: involvement and use. Table 18 shows the standardized canonical correlation coefficient for the two functions for each ATE, CETP, and LSC. Unlike the results produced when all programs were considered together, the results for each program individually shows that only the first function for each program is statistically significant.

Table 18.

Canonical Correlations by Program: Function 1

NSF Program	Canonical Correlations	Overlapping Variance %	Significance
ATE	.548	26	With both canonical functions included, $\lambda^2(4) = 56.68$, $p<.001$, and with the first function removed Function 2 was not statistically significant.
CETP	.760	58	With both canonical functions included, $\lambda^2(4) = 39.21$, $p<.001$, and with the first function removed Function 2 was not statistically significant.
LSC	.464	22	With both canonical functions included, $\lambda^2(4) = 14.17$, $p=.007$, and with the first function removed Function 2 was not statistically significant.

To get a better sense of how the individual programs may have differed from the aggregate results already discussed, the three different sets of coefficients produced by CCA will be discussed next. Table 19 shows the canonical coefficients, structure

coefficients, and the index coefficients for the first function for all three programs. The coefficients for the second function will not be discussed because they were not statistically significant, making the results discussed here different than the aggregate results.

Table 19.

Function 1 Coefficients by Program

Coefficients by Program	Canonical Coefficients Function 1	Structure Coefficients Function 1	Index Coefficients Function 1
ATE			
Involvement Factors			
Planning	.734	.977	.535
Implementation	.323	.874	.479
Use/Influence Factors			
Knowledge, understanding, or skills	1.140	.996	.546
Belief in the value of evaluation	-.172	.785	.430
CETP			
Involvement Factors			
Planning	.699	.954	.725
Implementation	.395	.846	.643
Use/Influence Factors			
Knowledge, understanding, or skills	.968	1.000	.760
Belief in the value of evaluation	.040	.794	.603
LSC			
Involvement Factors			
Planning	-.041	.156	.072
Implementation	1.000	.999	.464
Use/Influence Factors			
Knowledge, understanding, or skills	1.230	.904	.420
Belief in the value of evaluation	-.536	.200	.093

The canonical coefficients allow us to determine the relative influence or strength of the individual involvement and use factors to the overall relationship between the involvement and use. When the programs are examined separately, we see the findings diverge some from the aggregate results already reported.

ATE and CETP results mirror each other and share the same differences with the aggregate results and will be discussed first. Function1seems to describe a strong relationship between the planning factor and the knowledge, understanding, or skill factor. This differs from the aggregate results where implementation (an involvement factor) weighed more heavily in the relationship with knowledge, understanding, or skills (a use factor).

While the canonical coefficient allows us to begin describing the strength of the relationship between involvement and use, the squared structure coefficient allows us to speak more detail about the natures of the factors within the construct to which they belong, and the squared index coefficient allows us to look across the constructs. As such, we see structure coefficients for the involvement factors and use factors are in a relatively narrow range and can interpret this to mean that both involvement factors contribute strongly to the construct of involvement and both use factors contribute strongly to the construct of use. Similarly, we see index coefficients that are relatively close within and across factors and as such, we can say that both involvement factors moderately contribute to explaining the idea of use and both use factors moderately contribute to explaining the idea of involvement.

LSC results differ from the other two programs in substantive ways and will be discussed next. Much like the results in the analysis of all three programs in aggregate, the canonical coefficients seem to indicate that the involvement factor implementation and the use factor knowledge, understanding, or skills are most related. Unlike ATE and CETP where all factors were more balanced, in LSC we see that the involvement factor implementation contributes the most to the definition of the involvement construct and the knowledge, understanding, or skill factor contributes the most to defining the use construct. We also see that implementation contributes the most to explaining involvement's relationship with use; whereas knowledge, understanding, or skill contributes the most to explaining use's relationship with involvement.

Interview Findings

This study used information from 20 interviews with evaluators or principal investigators from projects involved in one of the three multi-site evaluations as a way to confirm or dispute some of the findings of the CCA and to add description. The following section will discuss these interviews in more detail.

First, all verbatim transcripts of the interviews were reviewed in detail. After that, each transcript was coded with a focus on describing or classifying emergent ideas conveyed in the interview. After all 20 transcripts had been coded the first time, the themes that emerged were compared to the factors for involvement and use in order to identify similarities and differences. The transcripts were reviewed and coded a second time in such a way as to uncover the more nuanced thoughts presented in the data (as when an interviewee discussed knowledge, what kind of knowledge was she referencing?

Or in what specific ways did interviewees describe findings use?). The relationship between Toal's involvement factors and Johnson's use factors and the themes that emerged from the interviews are shown in Table 20.

Table 20.

Interview Themes

Factors	Themes
Involvement-Related Themes	
Planning*	• Planned data collection for multi-site evaluation • Involved in development of multi-site evaluation plan
Involvement*	• Involved in instrument development for multi-site evaluation • Attended meetings/training/conference • Collected data • Reported data to multi-site evaluation
Motivation for Involvement	• For data-informed project decision making • Sense of professional "responsibility" • Prior experience with those involved • It was a condition of funding/program requirement
Use/Influence-Related Themes	
Increased Knowledge, Skill, or Understanding*	• Of my own project • Of program or NSF • Of STEM education or STEM fields • Of evaluation • As result of networking/collaboration with others involved in multi-site evaluation
Increased belief in the importance of evaluation*	• Changed beliefs about value of evaluation • Changed beliefs about what "evidence" is credible in determining project worth • Changed belief about their own role in contributing to evaluation
Indications of future use*	• Instrument use • Findings use • Use of elements of multi-site evaluation design in future evaluation planning
Influence	• Data collected across many projects allows for higher level "advocacy" or policy influence • Data collected across projects may have limited local project use • Influence on "the field" of STEM education or evaluation

* Indicates a theme that emerged in the interviews that is consistent with the factors identified by Toal and Johnson.

Involvement: Planning

The interviews described a wide range of involvement in planning of the multi-site evaluation—from no real involvement in planning the multi-site evaluation to active participation in instrument development or evaluation design. More specifically, interviewees all commented on their involvement in traditional evaluation planning activities. Some made specific reference to having no formal involvement in any multi-site evaluation planning activities. Others talked about being passively involved in planning when they were asked for input about the evaluation design or instruments. Some of these interviewees described specific ways in which their input informed some aspect of the final evaluation plan while others could not note how their input influenced the evaluation plan or instruments. Finally, others described more active participation in instrument development or evaluation design for the multi-site evaluation. This occurred primarily to local project evaluators or principal investigators who were called on to serve on an advisory board or planning committee for the multi-site evaluation.

Those who reported active participation in evaluation planning were involved in three traditional evaluation planning tasks: instrument development, evaluation design (sampling, methods, analysis, etc.), or reporting/dissemination of preliminary or final findings. Remember that the three multi-site evaluations fall at different points along the Cousins and Whitmore (1998) continuum regarding depth of participation. The ATE program provided fewer opportunities for local projects to be involved in evaluation planning, the LSC program mandated project participation, and the CETP program offered project involvement opportunities for local projects to collaboratively plan

aspects of the multi-site evaluation. These differing opportunities for involvement were evident in the interview responses. Most respondents involved in the ATE multi-site evaluation reported little or no involvement in evaluation planning activities. LSC respondents indicated participation in required activities and generally referred to their involvement as "*time-consuming*" compliance. Most comments indicated that PI meetings provided an opportunity for the multi-site evaluation PI to conduct member-checking of instruments, evaluation plans, or preliminary findings. Many LSC interviewees questioned the extent to which their input resulted in changes to the evaluation plan.

By contrast, respondents involved with the CETP program indicated a higher level of planned participation and collaboration in the development of instruments and plans. In addition, these interviewees expressed satisfaction that their input was considered and influenced the multi-site evaluation plans. Most CETP interviewees provided a specific example of an issue related to the instrument or data collection plans that they raised at meetings with the multi-site evaluation PI that resulted in changes to the instruments or evaluation plans.

The difference in the depth of involvement seems to come with consequences as well. Particularly those who reported their involvement as an obligation or as required expressed ambivalence about the multi-site evaluation planning process or complained of the cost or the amount of time required of them in the process. Examples of quotations from the interviews are found in Table 21 below.

Table 21.

Sample Quotations about Involvement in Multi-site Evaluation Planning

Activity	Sample quotations
Attended meetings	"When the grant would hold PI meetings…the PI would speak at those meetings. We'd have breakout groups [where we] may be asked for input."
	"We actually participated with a number of key people to begin to develop instruments. We carved out the vision for what the core evaluation would be and what it would look like and what kind of data we might collect...and how data collection could proceed in somewhat the same way in all of [the sites], and yet be tailored to each specific region or state or urban area."
	"We did what we were supposed to do and we went to the PI meetings. They would ask us to evaluate some of the stuff and to examine the instruments."
Instrument development	"I was contacted by Western Michigan to review the questionnaire, the survey, and provide feedback to them as to whether the questions were appropriate."
	"Actually, I think the first couple of years we were helping to make up the surveys."
	"I'm sure I commented on questions here and there. Probably more on the rating scale and the answers."

Involvement: Implementation

As might be expected based on Burke's (1998) key decision points that represent opportunities for meaningful stakeholder participation, the interviews highlighted a strong connection between the local projects and implementation of the multi-site evaluation. Interviewees reported involvement in the implementation of the multi-site evaluations in the following ways: a variety of data collection activities, contributing to reports, participating in communication/dissemination activities, and attending meetings. Conducting site visits, using observation protocols, and administering surveys designed

by the multi-site evaluation within their local projects were commonly described as involvement in data collection during evaluation implementation by most of the interviewees. For some local projects, the multi-site evaluation was implemented after local plans for evaluation had been determined. In these instances, the frustration of the local evaluators or PI is evident in the tone of their comments:

> …when we began [the multi-site evaluation] wasn't on the table. [Our project] had an evaluation with a budget…and no time was allocated for the national evaluation because it didn't exist. That was time put on top of what I needed to do. And I resented that because it took a good amount of time.

In other instances, interviewees described ways in which they summarized locally collected data, reported data, and in some instances verified data to the multi-site evaluation:

> I was the one supplying the data …or the explanation of some of the data that we submitted [to the multi-site evaluation].

> We reported back to the program evaluation, for example, all those activities we did in 1 year.

Again, the differences in the planned depth of participation within the multi-site evaluations were evident. Interviewees associated with ATE program spoke mostly about providing data and complying with data requests while interviewees associated with LSC and CETP discussed more opportunity for involvement beyond responding to directions from the national evaluation.

> …there were no invitations to be more involved in that there were no opportunities to meet with the [PIs of the multi-site evaluation] to talk about data use.—ATE interviewee speaking about involvement.

> I think that when we were asked, we felt that we were all heard, all listened to…it seemed like a much larger task than we could handle (laughs) but there were clearly things that had to be cut out or changed

> depending on the [local project].—CETP interviewee speaking about involvement.

Others described ways in which they were involved in reporting or disseminating findings:

> [After a site visit] we would write up our own independent findings and then had the opportunity to share them with others involved.
>
> [The multi-site evaluation] helped us to assemble our results and get them into a form where we could share them with others.
>
> The meetings allowed me to communicate and have further contact with other projects that are very similar to ours and we could share information and some research techniques and ideas.
>
> The same data we provided to the national evaluation, we used it in reporting to our advisory council.

Comments from interviewees indicated that several strategies employed by the multi-site evaluators to involve the local project PIs or evaluators throughout the entire evaluation fostered a greater sense of involvement. The strategies that garnered the greatest level of buy-in and participation included actively soliciting input at all stages of the evaluation and facilitating or creating opportunities for the local project people to network and collaborate with one another around the evaluation work.

One finding of the interviews differed from Burke's key decision points regarding stakeholder involvement. Interviewees repeatedly discussed their participation in or attendance at meetings related to the multi-site evaluation as involvement:

> …and also, the only way I was involved was at the PI meetings.
>
> They also provided training at the beginning before we started and then annual meetings that we went to talk to the other evaluators and get updates from them…
>
> We would meet annually and we would discuss how things were going.

> We participated in the meetings because we were eager to find out about our impact, of course, and to begin to think about how to improve what we were doing because we wanted it to have a lasting effect.

It appeared that interviewee comments about the involvement in meetings referred to all types of meetings—face-to-face meetings, conference calls, webinars, and other asynchronous electronic exchanges. While most interviewees valued all types of meetings, the most favorable comments referred to face-to-face meetings. Interviewees expressed that one of the valued outcomes of all types of meetings was the chance to network and collaborate with people involved in other projects, and this was especially true for those associated with LSC and CETP.

Use/Influence: Knowledge, Skill and Understanding

The interviews were rich with examples of how local project evaluators and PIs gained increased knowledge, skills, or understanding because of the multi-site evaluation. Interviewees from each of the three programs indicated that they increased knowledge of their own project or of the national program as a result of involvement in the multi-site evaluation. Examples from their comments focused on increased knowledge or understanding of what constituted indicators of high quality STEM education, improved knowledge of processes or activities that contributed to program operation, more concrete knowledge of their program reach, processes, or outcomes, and an improved understanding of how their local project contributed to the national program goals/achievements. The interviewees described some specific uses for this new knowledge or understanding as well. They indicated that it was used for their own professional development, for program planning and improvement, or to inform or

persuade others interested in their local project. Samples of related quotations include the following:

> The most important thing was it gave me a sense of what NSF was really looking for in the program…a sense of what the metrics were.
>
> It helped [our team] to be more self-reflective and regularly review our plans.
>
> The classroom observation protocol impacted me personally. It was a nice way to think about that not all inquiry was good, and that it had to be high-quality inquiry, not just activity for activity's sake.
>
> The fact that [the evaluation] had research coming out that supported what we did here, that was important with the powers-that-be here in our district.
>
> [I gained understanding] that you need to have some kind of evidence to show anybody whether something is working or not. And it's not enough for the project to know about what's going on but I think that the world should know too.
>
> Basically, I think it made us look [at our project] more systematically. I think in the fact of knowing that the data I was collecting was being done in a way to be compiled with other people's data and seeing that accomplished gave us a sense of the importance of what we were doing and helped us get past the frustrations of not succeeding as well on the local level because of all the garbage of the school system.

In addition to gaining program or project knowledge because of the multi-site evaluation, a strong theme that was evident in interviewees' comments related to increased knowledge of evaluation. Evaluators and PIs involved in local projects related to the three multi-site program evaluations described how they developed a more thorough understanding of the role or possible uses of evaluation. Several interviewees described how the multi-site evaluation prompted self-reflection or a change in discourse among project staff that resulted in changes to local evaluation plans as well.

> It helped us rethink or brought our team together to think about our evaluation.
>
> This was my first evaluation job and I learned a lot.
>
> I think it had significant impact on how we went about doing evaluation of the project.

There was one theme that emerged from the interviews that was not supported in earlier literature about evaluation use that shaped instrument development for the Beyond Evaluation Use study. Many interviewees referred to knowledge that resulted from the opportunity to network or collaborate with other project sites within the multi-site evaluations or from the interaction with the veteran multi-site evaluators. Some of the described gain in knowledge seemed to result from peer discussions at PI meetings or training that was a part of the multi-site evaluations. One interviewee described knowledge gains about both project operations and evaluation because of the relationships with those involved in other projects as *"very useful."* Other gains in knowledge seemed linked to evaluation techniques that were modeled by the multi-site evaluators or resulted from direct feedback from the multi-site evaluator to local PIs or evaluators about local evaluation activities.

> So I think for me in particular it was a really good experience to be part of the collaborative planning of those evaluation instruments. Like right now, I'm working on a grant proposal with them and it's had a lasting effect.
>
> The training and the discussions within the group at the training was particularly useful to me.
>
> I could see what other people did on their campuses…so that larger perspective was useful and I'm sure it influenced our future projects.

Beyond changes in knowledge, the interviews also revealed many references to skill development that occurred for PIs or evaluators of the local projects within the

multi-site evaluation. Interviewees referred to gaining skills in evaluation as a result of their involvement in the multi-site evaluation. More specifically, a number of interviews revealed that they learned to use new evaluation activities, including the development and use of logic models, the appropriate use of a variety of data collection techniques, and the development or improvement of skills in evaluation planning.

> [This multi-site evaluation] really grounded us in what we had to produce and things we had to think about. I really think it helped us plan evaluation activities locally.
>
> I enjoyed learning about how the whole process of how the instruments were developed…how the whole process was put together.
>
> They [the multi-site evaluators] educated us about some components of evaluation that we were relatively unfamiliar with, things like logic models, for example. We went on to use them in our own project

Skills in developing evaluation instruments and tools were also referenced by a number of the interviewees. The connection between this skill development and the interactive nature of multi-site evaluations was also evident as several interviewees noted how they used tools and resources from the multi-site evaluation to teach or support others involved with their project or other future projects. While many interviewees noted that they gained new skills because of the act of carrying out activities associated with the multi-site evaluation, some felt they gained new skills by trying to replicate techniques modeled by the multi-site evaluators.

> The skills I learned I now pass on—interview techniques, ways of documenting interviews, ways to look across interviews—making similar judgments and using some of our coding across interviews.
>
> We used the instruments to develop teacher training within our project…we ended up using the instruments as a way of helping teachers understand what a good lesson looked like.

Use/Influence: Changes in beliefs about evaluation

While not as richly described in the interviews as the changes in knowledge, skill, or understanding, interviewees' comments also reflected changes in their beliefs about evaluation as a result of their participation in the multi-site evaluation. The changes described fell into one of three broad categories of belief changes: changes in belief about the value or importance of evaluation; changes in their beliefs about what makes credible evidence of a project's merit; or changes in their beliefs about their role in evaluation. Table 22 shows how sample quotations aligned with these three broad themes related to changed beliefs.

Table 22.

Sample Quotations Regarding Use/Influence—Changes in Beliefs

Changes in beliefs about:	Sample quotations
The value or importance of evaluation	"[Through the multi-site evaluation meetings] The opportunities helped my understanding [of evaluation] really evolve over time. It's (evaluation) is an important part now."
	"I particularly like to use [evaluation processes or findings] with projects just getting started."
The role of credible evidence to determine a project's merit	"[The multi-site evaluation helped us see that] In a sense you need to have some kind of evidence to show anybody whether something is working or not."
	"Certainly all the data that I collected was compiled with everybody else's data and though the process was grueling, it helped us to see where we had succeeded, where we had failed, what we had problems with and what we could do differently in subsequent years."
	"The fact that the multi-site evaluation had research coming out that supported what we did is important. [Project stakeholders] might say, 'Maybe you should think about doing this or that.' But we could say, 'this is what the research is showing.' And so I think having research data and referring to it helps the vision."
Their role in evaluation	"I've actually gotten people to look at [evaluation] and think about how they can use it in their projects."
	"I think the was that it was initially handled—how all the projects were brought together to plan—was important to the success of the evaluation process as well as keeping [project evaluators or PIs] involved and committed to the process."

Use/Influence: Indications of Future Use of Knowledge/Skills Gained

As described earlier in this chapter, a subset of the survey respondents answered survey questions about their future use of knowledge or skills gained through their involvement in the multi-site evaluation. Given the already reduced sample that provided information on this topic, the interviews offered an even more limited look at the idea of future use. Still, comments from the interviews indicated that project evaluators or PIs

did use knowledge or skills from the multi-site evaluation in future work. The interview evidence pointed to three categories of future use that were common among interviewees: future use of new knowledge related to evaluation planning/design, future use of instruments from the multi-site evaluation, or future use of multi-site evaluation findings.

When interviewees referred to future use of new knowledge/skills related to evaluation planning/design, they often spoke of how successful experiences in the multi-site evaluation planning prompted them to use or modify elements of the evaluation design in future projects. One person noted, for example, "[We used our multi-site experience] to guide some of the evaluation questions we asked in other projects in subsequent years."

There were several instances where interviewees noted future use of instruments from the multi-site evaluation. Uses of the instruments seemed to be focused on training of project stakeholders or modification of the instruments for data collection in other projects.

> I went on to serve a new role in a different project but the first thing I did was to go to the program website and pull down the resources from the evaluation to plan my new work.
>
> I've actually gotten people to look at it and think about how they can use [the instruments] in developing their own evaluation [for a different project].
>
> I went on to be an evaluator in a couple more projects and I started some things in these new projects in terms of surveys [based on the surveys from the multi-site evaluation].
>
> I have used the classroom observation protocol in other work.
>
> I sort of modified their instrument and have used it to survey other teachers here.

Many interviewees noted how they used findings from the multi-site evaluation in future work. Their findings use focused on using findings to secure future grants, to design future projects, or to develop professional development in organizations.

> [The multi-site evaluation] findings were helpful in trying to acquire new funding…and saying OK, here's what worked across a lot of different projects.

> [The impact] is almost more after we were finished with the multi-site evaluation. That retrospective piece has impacted some of the professional development activities we are running at our university and some of the new grants we've written.

> The data our group was [sic] able to get and use has subsequently helped us develop new programs and get new grants.

Two ideas that emerged in interviews that did not fit into the research scheme regarding use had to do with the role of dissemination and its influence on use and the timing of the start of the multi-site evaluation in relation to the life cycle of the local project. Numerous interviewees noted missed opportunities and challenges in these two areas. For instance, interviewees from all three programs noted uncertainty about final evaluation findings or the whereabouts of final reports. In contrast, other interviewees from all three programs noted that the various ways the multi-site program evaluation used to disseminate information and findings throughout the evaluation encouraged local findings use (i.e., reports, booklets on a single theme, website data files, etc.). Several interviewees also noted that the feedback loop of data seemed to fall short of delivering findings to local projects:

> I think the findings got to the NSF leadership but not back to the college-level folks in the projects.

> We tried to…request the database of answers we provided so we could do other analyses on it but what we got back was so confusing that we couldn't figure it out.
>
> We met once a year during the first 2 years of the project to get updates and I assumed that we would have a similar meeting during the third year and by that time there would be a compilation and analysis of all this data that everyone was collecting—that we would all be able to draw conclusions from it, but the third year the meeting never existed.

In addition, the multi-site evaluations were not all planned when the multi-site programs were launched. As a result, participation in the multi-site evaluation was asked of projects at different times during the lifecycles of the local projects. Some interviewees were able to plan and launch their local project fully informed of their role and requirements for participation in the multi-site evaluation while others were asked to participate in multi-site evaluation activities when their project was well underway. As one might expect, this difference in the multi-site evaluations is evident in the conflicting comments of local evaluators and PIs involved in the three programs.

> The [multi-site evaluation] came to life after the fact and there was a lot of negative feeling by evaluators who were working with little guidance.
>
> These 5-year projects, with projects coming and going at different times, can see a lot of burnout as time goes on. This made gaining buy-in [to the multi-site evaluation] difficult.
>
> They gave us instrumentation that was pretty well done but it added to our workload because we already had other local data collection plans before the big program evaluation came about.
>
> I thought the whole evaluation was well done…it provided a framework, which was really pretty global, on all the topics you could possibly fit into an evaluation of this kind. It set us off on the right foot from the beginning.

In summary, the results of the canonical correlation analysis, suggest a number of assertions. First, the bivariate correlations within the involvement factor set were

moderate and within the use factors were moderate to strong, indicating that the factors were reasonable representations of involvement and use. In addition, bivariate correlations within the two factor sets were generally higher than the correlations across factor sets, supporting the idea that involvement and use are separate constructs.

The standardized canonical correlation showed us that there was a moderate correlation between involvement and use with almost one third of the variance shared between the two factor sets. A number of additional canonical coefficients illuminate more about the nature of the relationship. First, the standardized canonical coefficient allows to us identify the strongest relative influence of involvement on use. We see that the knowledge, understanding, and skill factor (related to use) and the implementation factor (related to involvement) contribute the most to the first canonical function and therefore, more importantly, account for the maximally correlated combination of factors in the two sets.

Next, the squared structure coefficients showed us that implementation contributes the most to defining the involvement construct and knowledge, understanding, or skill contributes the most to defining the use/influence construct and corroborates how these two factors make up Function1as the most highly correlated combination of factors.

Finally, the squared index coefficients allow us to understand how the involvement factors relate to the use factors and the use factors relate to the involvement factors. This study shows that implementation contributes slightly more than the planning factor to explaining the relationship to use/influence. Similarly, the knowledge,

understanding, and skills factor contributes more to explaining the relationship with involvement.

When examining the interview data, there was a greater density and diversity of comments about implementation compared to comments about planning as interviewees discussed matters related to involvement. Likewise when aligning comments related to use/influence, we find that comments related to knowledge, understanding and skills were more prevalent and diverse than comments about changes in beliefs or indications of future use. The interviews also provide insight into interviewees' motivation for involvement and offered insight into "influence" on a broader group as a form of use not often seen in the literature on but suggested in Kirkhart's model.

Chapter 5 will provide an explanation of how the results are relevant to the research questions and discuss the extent to which the results describe or explain the relationship between involvement in and use of evaluation in a multisite evaluation context.

CHAPTER 5

DISCUSSION AND IMPLICATIONS

While the relationship between participation in and use of evaluation is well documented in single-site, local program evaluations, little has been examined about the relationship in a multi-site evaluation context. Beyond that, little is known about the nature of the relationship between involvement and use among users that are not primary intended users who may be involved because of the culture or expectation of evaluation funders. It is important to note that this study did not seek to determine whether or not there is a clear distinction between use and influence or the role played by these multi-site evaluations in exercising influence in indirect or unintended ways as described by Kirkhart (2000) and others and has treated use/influence as synonymous during analysis and interpretation. This chapter will, however, provide an explanation of how the results outlined in the previous chapter are relevant to the research questions and discuss the extent to which the results describe or explain the relationship between involvement in and use of evaluation in a multisite evaluation context among secondary or unintended users. The evidence for the supporting research questions will be discussed first, beginning with the quantitative results, followed by the qualitative results, and will conclude with a discussion of how the findings converge with and diverge from the current literature. Then, there is a discussion of how the supporting questions answer the primary research question. Finally, this chapter will conclude with a discussion of implications for practice, implications for research and limitations of this study.

Supporting Question 1

> To what extent can evaluation use be explained by participation in multi-site evaluations?

Substantive Findings of CCA

The canonical analysis yielded two canonical functions that represent the strength of the relationship between involvement in and use of evaluation. The standardized canonical correlation coefficient for the first function was .558 (31% overlapping variance); the second was .13 (2% overlapping variance). With both canonical functions included, $\lambda^2 (4) = 103.22$, $p<.001$, and with the first function removed $\lambda^2 (1) = 4.79$, $p = .029$, making both functions statistically significant. This means that there is a statistically significant relationship between involvement in and use of evaluation in the multi-site settings that were involved in this study. While this is a valuable finding, it is important to understand practical significance, not just statistical significance. Other results delivered by canonical analysis allow us the needed insight into the practical importance and will be discussed next.

Supporting Question 2

> What contribution does any single factor related to involvement in evaluation make to explaining involvement? What contribution does any single factor related to use of evaluation make to explaining use?

Substantive Findings of CCA

The literature on participation or involvement offered a variety of frameworks and models that provide descriptive characteristics of participation. The three-dimensional framework of Cousins and Whitmore (1998) and Burke's key decision points (1998) provided structure to the conceptualization of this study in that the depth of participation

was examined as well as the ways in which project staff was involved in the various evaluation activities of the three multi-site evaluations.

The standardized structure coefficients from the canonical correlation analysis shed light on which specific factors identified by Toal contributed most to our understanding of the construct of involvement in multi-site evaluation. The results indicated that in Function 1, implementation (.96) was slightly more influential than planning (.81) in the explanation of the involvement construct and Function 2 was most strongly influenced by planning (.57) in the explanation of the involvement construct. Remember that CCA combines factors related to involvement and factors related to use into functions so as to maximally correlate the factors. When defining implementation activities in alignment with Burke's key decision points, these CCA results indicate that when project staffs are involved in developing instruments, collecting data, or reviewing/analyzing/interpreting data, the staffs are more likely to describe themselves as involved in the evaluation.

With regards to the construct of use/influence as defined by the factors identified by Johnson, the squared structure coefficients show that Function1was more influenced by the knowledge, understanding, or skills factor (.99) and Function 2 was strongly influenced by the belief in the importance of evaluation factor (.75). As defined in the EUG study, the knowledge, understanding, or skill factor included many ideas like knowledge/understanding of planning an evaluation, implementing an evaluation, communicating findings, STEM education, STEM education evaluation, or the local

project as well as increased skills in planning or implementing an evaluation, communicating findings, or work for the local project.

Substantive Findings of the Interviews

A majority of the interviewees confirmed the prominent role that involvement in implementation activities had on participants' perceptions of involvement. Many interviewees noted examples of their involvement in the multi-site evaluation as development of instruments or collection/reporting of data. Almost all interviewees noted dissemination activities as involvement. This is important because within the NSF context, dissemination is viewed as a type of involvement. However, NSF's emphasis on dissemination to the field varied across the four projects and over time. Far fewer interviewees described involvement as participation in evaluation planning activities.

The interviews were rich with examples of how local project evaluators and PIs gained increased knowledge, understanding, or skills because of the multi-site evaluation, supporting the CCA findings. Far fewer interviewees discussed examples of the belief in the importance of evaluation factor.

Convergence and Divergence of Findings

The findings of this study related to Supporting Question 2, regarding involvement in evaluation activities, generally align with the literature. Findings of this study confirm that the various ways participants' reported involvement in a multi-site evaluation aligned with tasks identified in the literature, including evaluation planning, instrument development, data collection, data review, or analysis and interpretation of data. In addition, outreach and dissemination are not the same as more direct involvement

in the evaluation processes, but do serve as a type of involvement in multi-site settings. Thus, given that project staff heavily defined their involvement as dissemination and outreach work within a multi-site evaluation, this is an important divergence from the literature. Similarly, the findings in this study confirm that in a multi-site context, the same four types of use already established in the literature exist: instrumental, conceptual, symbolic and process use.

These findings diverge from the literature in one important way. When referring back to the results of the 1996 and 2006 surveys of Canadian and American evaluators (Cousins et al., 1996; Fleischer & Christie, 2009), the tasks that evaluators most often reported involving stakeholders in were planning tasks (discussions that focus the evaluation, identifying evaluation team members, developing the evaluation plan, planning instruments or data collection) that make up the planning factor in this study. In a multi-site context however, the evaluation often serves the purpose of collecting standard data and information across a multiple sites in order to look across sites for the greater program impact and serving the funder as the primary intended user. These multi-site evaluations in turn, generally allow less opportunity for local project staff to have input in the planning phase of the multi-site program evaluation. The subjects of this study more often reported feeling they were participating or involved in the evaluation when they were engaged in tasks aligned with implementation of the evaluation. When members of the evaluation profession's leading association indicate frequent involvement of evaluands in the planning aspects of evaluative work and evaluands in the field report feeling most involved when they are working on evaluation tasks associated with

implementation, the discord may be have significant influence on use and may be important beyond the multi-site setting.

Supporting Question 3

> Do certain factors related to participation contribute more to explaining the construct of (all factors related to) use/influence?

Substantive Findings of CCA

Using the standardized index coefficients allows us to look at the extent to which a single involvement factor contributes to explaining the overall relationship with the construct of use. Index coefficients are more conservative than the structure coefficients, but the results of the CCA show the same factors contributing the most to explaining the relationship between involvement and use. The index coefficients for Function 1 show us that the implementation factor (.53) contributes slightly more than the planning factor (.45) in explaining the relationship with the construct of use/influence. In other words, we might expect to see use of evaluation processes or findings when multi-site evaluation project staff are involved in evaluation implementation-related activities like developing data collection instruments or processes; collecting data; reviewing, analyzing and/or interpreting data; writing reports; or presenting evaluation findings.

Similarly, the index coefficients in Function 1 also show us that the knowledge, understanding or skills factor (.55) contributes the most in explaining the nature of the relationship with involvement. This tells us that a project participant is more likely to gain knowledge, understanding, and skills if they have been involved in a multi-site evaluation than to experience a change in beliefs about the value of evaluation.

Substantive Findings of the Interviews

Making a connection between the CCA findings and the interviews related to question RQ2b is not as direct as linking specific stories found in the interview to the statistics that indicate which factors combine to account for the greatest amount of shared variance. However, there were themes found within the interviews that allow us to make connections between an individual factor related to involvement and the overarching construct of use.

First, many interviewees associated with LSC or CETP noted involvement in instrument development or data collection using the instruments developed for the multi-site evaluation. Upon examination of these respondents' comments about use later in the interview, there were also many indications of process and findings use from the act of gathering those data and from the reports generated with those data. This supports the CCA findings linking involvement in implementation tasks with use.

Another theme within the interviews that sheds light on the relationship between the involvement factors and use is related to the effect of local project control on involvement in the evaluation on use of processes and outcomes. One program in this study "required" local project participation in the multi-site evaluation activities which, of course, resulted in involvement, but this enforced involvement didn't necessarily result in more use of processes or outcomes. Interviewees from the program that required their participation often spoke of frustration with regards to the amount of time the required activities took away from their local work. In addition, many of these same interviewees noted that the scale and size of the program meant that data collected and findings

generated were not focused on indicators of interest to them or useful for their local project.

There is one additional theme that seems to fit here—the role of collaboration and networking in facilitating use. Many efforts in education reform are built on the belief that involvement in a community of practice or network creates the sense that people share in the benefit of the group's efforts. As such, the networking allows for the identification of different strengths or capacities and the feeling of being part of a team. Team members feel more obligations to work with the team and feel part of what the team produces. In the context of this study, this could contribute to more use of the processes and products of the multi-site evaluations. All but one of the evaluation projects showed clear evidence of the development of community and networking among the individual projects, and the exception still had at least some evidence of networking. Most often, use as it related to this theme of networking was process use.

Convergence and Divergence of Findings

Although not all evaluation theorists agree that process use constitutes a unique category of use (Alkin & Taut, 2003; Johnson, 1998), the findings of this research support the idea that process use is distinct. Project staff often reported that the act of meeting with people from other projects or collecting data taught them something about their own project or caused them to think differently about their project plans, activities, or procedures.

The findings related to local project control lend credibility to the model proposed by Lawrenz and Huffman (2003), which called for multi-site evaluations to be negotiated

yet centralized. The negotiation of a process that collects data from local sites and aggregates it up to the larger program (centralized) while collaboratively designing the evaluation in such a way as to integrate the local needs with those of the larger multi-site program might have removed some barriers created by the required participation noted by some interviewed for this study.

However, these findings also diverge from a number of theories and assumptions currently held about participatory evaluation in single site settings. First, the very definition of "participatory" is transformed in these multi-site contexts from what has been described as participatory in a single site setting (Burke, 1998; Cousins & Whitmore, 1998). The program in this study where local projects felt their involvement was mandatory would not meet the characteristics of involvement outlined by King and Stevahn (2002) as participatory, yet participants connected to that program noted numerous examples of their own involvement and referred to this involvement as participatory. Similarly, interviewees in all programs described "attending meetings" as one form of their involvement—certainly not aligned with earlier notions of participatory evaluation in the literature.

In addition, the perception of "required" participation in one of the programs studied challenges another idea advanced by Mark and Shotland (1985). Their work discussed the role of power and legitimacy in the selection of involved stakeholders and noted that those stakeholders in a high-powered or influential position who are involved in the evaluation will increase likelihood of enhanced use. In the case of the programs studied here, the project PIs, acting as a stakeholder on behalf of their local project, were

often the involved party in evaluation activities and decision making. As project PIs, these individuals were certainly in a position to make decisions based on the evaluation findings, but those PIs representing the projects in the group where they perceived their participation to be required reported very limited use of the processes or findings of the multi-site evaluation because of the disconnect between their role in planning and decision making about indicators and measures examined for the multi-site evaluation. More often, any use described by this subgroup was, as King (1988) called it, signaling or compliance with requirements of funders. Finally, the role that meetings and collaboration played in fostering use fell outside Burke's key decision points and seemed unique to the multi-site context.

Research Question 1

> What is the nature of the relationship between participation in and use of evaluations in a multi-site evaluation setting?

Substantive Findings of CCA

The bivariate correlations indicate that the factors within the involvement and the factors within use adequately reflect the two separate constructs. In addition, the across pairs correlations support the idea that two constructs are independent of each other.

The CCA results allow us to describe a number of aspects of the relationship between involvement in and use of evaluation in a multi-site setting. In Function 1, the linear combination of the factors representing implementation activities and the factors representing use show that involvement in implementation and increased knowledge, understanding, or skills are more prominently related than the other factors and account for almost one third (31%) of the shared variance for the factors examined in this study.

Considerably smaller but still significant, the involvement factor representing planning and the use factor representing an increased belief in the importance of evaluation are more prominently related than the other factors and accounted for 2% of the covariance. In essence, this positive relationship tells us that one third of the time, when involvement increases, use increases or as use increase involvement increases—especially when project participants are involved in activities related to evaluation implementation (developing instruments, collecting data, reviewing, analyzing, or interpreting data, or reporting findings) and the kind of use that is likely to be most related is conceptual use (increased knowledge, understanding or skills). This is important in building our understanding of the nature of the relationship between involvement and use.

Not only can we judge the direction of the relationship, we can also judge the strength of the relationship from the standardized canonical coefficient. The value of this coefficient can only range between 0 and the number of variables in the smallest set (in this case 2). In this study, we see a moderate strength in Function 1 between implementation (.74) and knowledge, understanding, or skill (1.17) and slightly greater strength in Function 2 between the factors increased belief in value of evaluation (1.54) and planning (1.20). This is to say that in the function that represents the strongest correlated factors shows the greatest covariance between the involvement factor implementation and the use factor knowledge, understanding or skills.

Substantive Findings of the Interviews

The positive and moderately strong relationship between involvement and use in multi-site settings was supported in the interview comments, especially as interviewees

talked about their motivation for getting or remaining involved in the multi-site evaluation activities. For instance, the interviewees who noted some of the highest levels of use also reported a number of more explicit strategies employed by the multi-site evaluator to gain and support their continued involvement in planning or implementation tasks.

Summary of CCA Findings by Program

To determine consistency in findings between the full set of all three programs and the findings when the data file was split by program, CCA was conducted on each of the three programs in isolation. The process had the potential to further illuminate the nature of the relationship between involvement and use because the three programs: ATE, CETP, and LSC all operated from a different place along the Cousins and Whitmore's (1998) depth of involvement dimension. The effect of the reduced sample size became evident in the results from the separate analyses. The CCA was only able to demonstrate statistical significance of one of the two canonical functions produced by the procedure. Each program demonstrated a moderate correlation between involvement in and use of evaluation with CETP showing a strong correlation (.76). When examining the role of the individual factors in describing the relationship between involvement in and use of evaluation, the ATE and CETP results diverged from the results of the aggregate analysis. While the aggregate analysis showed the greatest relationship between the involvement factor implementation and the use factor understanding, knowledge, or skills, ATE and CETP results showed the greatest relationship between the involvement factor planning and the use factor knowledge, understanding, or skills. Initially, it might

seem unusual for ATE (with the lesser depth of participation) and CETP (with the greatest depth of participation) among the programs to perform similarly however one characteristic the two programs shared was the voluntary nature of involvement. LSC was the only program that mandated local project involvement in certain evaluation activities though they were not involved in the depth of activities with which those in CETP were engaged. In the separate results for LSC, we again see that the involvement factor implementation and the use factor knowledge, understanding, or skill play the greatest role in describing the relationship between involvement in and use of evaluation.

Summary of the Findings

This study examined the intersection of involvement in and use of evaluation in multi-site evaluations. It determined that the strength of the relationship between participation and use is moderate to strong in the program evaluations of interest and positively correlated.

More specifically, we see a moderately strong relationship between project staff who were involved in implementation activities like developing instruments; collecting data; reviewing, analyzing or interpreting data; or reporting findings and reported evaluation use in the form of increased knowledge, understanding or skills related to evaluation when examining the CCA and interview results together.

In addition, this study also identified a number of ways in which the multi-site setting diverged from the single site evaluation context—each difference bringing with it changes to the relationship between involvement and use. For example, the selection of stakeholders who are involved in the evaluation differs in that local projects serve as a

representative stakeholder to the multi-site project on behalf of all their local project stakeholders. In addition, the NSF context values the role that dissemination and outreach of lessons learned plays in "growing the field." As such, many funding recipients are required to attend networking and collaboration meetings within the programs. While these meetings were not explicitly evaluation activities, they played substantive roles in describing how project team members felt involved in this research.

Limitations

There are several methodological limitations of this study related to the use of surveys. First, the data in this study were self-reported, introducing the possibility that the data were influenced by the desire on the part of the program PIs or evaluators to appear in a particular manner. Although confidentiality was assured to participants in the current study, the Beyond Evaluation Use research project that collected the data used in this study was headed by a well known NSF evaluator whom respondents may have felt could exercise influence over their current or future NSF-funded work. Second, despite the Beyond Evaluation Use research team's efforts to clarify their interest in responses related to the multi-site program evaluation, respondents may have responded from the perspective of their own local project evaluation. Finally, memory problems may have also affected respondents' abilities to accurately reply to all survey items and interview questions. For some of the respondents, 8 years may have passed between the time of their work on the project and the completion of the survey/interview used in this study.

In addition, a few study limitations involve statistical issues. Structure coefficients, used to describe which factors contributed the most to defining its own

construct, may be subject to considerable variability from one sample to another. This variability suggests that the coefficients, and hence the relationships ascribed to them, may be sample-specific, resulting from chance or extraneous factors (Lambert & Durand, 1975). While this study did not employ a sample but rather relied on the entire population of the four NSF-funded programs of interest, a researcher still must be cautious when interpreting canonical relationships, particularly with regard to the external validity of the findings. Next, the size of the population was not enough to conduct an analysis of the full model including all three factors related to use/influence that were identified by Johnson (2008). As such, it is unknown how the third factor related to use (future use) may change the linear relationships within the canonical functions had the number of responses in each cell been adequate to examine the full model. Lastly, the data from this current study were subjected to multiple statistical tests and comparisons which may have increased the likelihood of Type I errors, despite efforts to minimize this through the application of statistical correction techniques.

Implications for Practice

A number of the findings from this study provide evaluation practitioners an opportunity to increase the likelihood of use by leveraging ways in which they involve project staff in evaluation tasks. Based on the findings of this study, there appear to be a number of guiding questions an evaluator might ask him/herself in planning a multi-site evaluation if enhanced use of the processes or findings is important:

1. What role will I ask local project staff to play in different evaluation activities? (i.e. planning the evaluation, designing data collection

instruments, collecting data, analyzing data, reporting or disseminating findings)

2. Are there data concerns that arise because of differences across the multi-site locations?
3. Are all projects (and project staff) at a similar place in the program life cycle?
4. How can I use communication to foster and support buy-in and involvement?
5. What can I do as an evaluator to moderate tensions that will exist between the multi-site program and the local project?

The choices made in answer to all these questions may contribute significantly to the likelihood of enhanced use based on the findings of this study and the work of the Beyond Evaluation Use study (Lawrenz & King, 2009). Findings from this dissertation research suggest that multi-site evaluators may enhance participants' involvement in a multi-site evaluation by leveraging the role they ask participants to play in implementation tasks (like instrument development, data collection, data review, or analysis and interpretation of data) over the role they ask participants to play in planning activities. However, it is important for multi-site evaluators to exercise caution and not expect too much involvement of local project staff in order to avoid creating feelings of excessive burden on local project staff. In addition, a multi-site evaluator should not underestimate the need for ongoing communication and opportunities for collaboration among participants as a way to foster involvement.

The findings related to use/influence suggest that multi-site evaluators can enhance use/influence of evaluations by designing multi-site evaluations that explicitly plan for any of the recognized forms of use (instrumental, conceptual, symbolic and process use) from the beginning. This suggests that an important role of the multi-site evaluator is to plan for ways in which the multi-site evaluation delivers something of use to each local project through thoughtful design and the inclusion of deliberate tasks that build evaluation capacity related to use by local project staff. Numerous times, participants in this study provided evidence of how the one multi-site evaluation with which they were involved was useful in providing them with tools to use in future evaluation work, in helping them shift their thinking about evaluation or STEM education, in making operating decisions for their project, or in persuading other key stakeholders. In addition, planning ongoing opportunities for collaboration among participants is an important way to enhance use in ways that may be unique to a multi-site context.

Beyond implications for evaluation practitioners, this study offers guidance on roles funders may play to enhance use of multi-site evaluations as well. Funders may play many roles and make many decisions that influence when and in what ways people get involved in evaluation and whether and in what ways people use evaluation. Some critical roles they play can be enhanced by this research. First, funders often have opportunity to influence policy in their daily work. By funding grantees to educate decision-makers about evaluation use and by rewarding and encouraging the implementation of effective practices, evaluation involvement and use within multi-site

evaluation might be greatly enhanced. Second, funders often provide a conduit for fostering networks or collaboration and for educating the field. If funders were to use the knowledge that involvement in the actual implementation of evaluation activities is more highly valued by project staff, the funder can structure technical assistance and professional meetings in such a way as to build local project staff capacity to implement evaluation activities rather than allocate lots of time and expect high levels of local involvement in planning. As important as facilitating project involvement in evaluation, funders can often play a critical role in providing distinction and guidance on times when involvement may not be appropriate for the individuals involved or the program as a whole.

Given that this study highlighted how local project staff was more likely to report using knowledge, understanding, or skill that resulted from their involvement in the evaluation, a funder should plan explicit professional development and capacity building activities into the program's operation. One under-recognized role that funders play in evaluation practice is in strengthening individual and organizational skill and knowledge about evaluation.

A number of data quality concerns were expressed by participants in this study. In determining standard measures or indicators for the multi-site program, funders should also take great care in defining the measures and in helping local project sites collect, save and report this information to the multi-site evaluators. In addition, if a funder chooses to require participation in a multi-site evaluation, it should be an explicit expectation before a project becomes operational and should allow for the local project to

also collect and use data of interest to them for local project improvement and project promotion. In addition, funders who operate programs where local projects enter and exit the program at different times should be aware of the conflicts this life cycle diversity produces among the local project sites and plan deliberate technical assistance to mitigate those difficulties. In seizing the opportunity afforded to funders by the nature of their role in program operations and evaluation implementation, funders can contribute considerably to building program capacity to engage in evaluative activities and in using evaluation processes and findings.

Implications for Research

This work was the first to use quantitative methods to examine the nature of the relationship between involvement and use of multi-site evaluations. As such, the themes and findings of this research provide fruitful avenues for future research. For instance, replicating this work in fields beyond STEM education and with a larger sample might allow for examination of the full model of both involvement factors and all three use/influence factors. In addition, a lot may be gained in our understanding of the relationship between involvement and use if we examined the intended users of the evaluation (in this case NSF).

Issues raised in this work might be examined more effectively by employing quantitative path analytics, which would allow researchers to describe the directed dependencies among a set of variables or allow researchers to test theoretical propositions about cause and effect without manipulating variables.

In addition, data issues present in this study that came from a lack of consistency in perceived involvement and use made measuring involvement and use challenging. In response, continued efforts to refine the involvement and use scales in different contexts could also enhance validity and reliability. However, improving methodology and measurement alone may not be enough as any investigation would likely to be substantially affected by the nature of the evaluation and the characteristics of the individuals involved.

Yet, despite the many unanswered questions that remain about the relationship between involvement in and use of evaluation, this study contributed to knowledge in the field in several ways. The findings shed light on the many leverage points between the degrees of involvement and levels of use that are reasonable from multisite evaluations in which stakeholders are secondary audiences who are involved in the multi-site evaluation a variety of ways. While we still do not know the direction of the influence between involvement and use, we can empirically say a relationship exists. By advancing our knowledge of the nature of the relationship between involvement and use this study provides practitioners, researchers, and program funders alike with reasoned assertions to strive for improvement in the way multi-site program evaluations are designed and operated.

REFERENCES

Alkin, M. & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, *29*, 1–12.

Alkin, M., Hofstetter, C. & Ai, X. (1998). Stakeholder concepts. *Advances in Educational Productivity*, *7*, 87–113.

Amo, C., & Cousins, B. J. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. *New Directions for Evaluation*, *116*, 5–26.

Ayers, T.D. (1987). Stakeholders as partners in evaluation: A stakeholder-collaborative approach. *Evaluation and Program Planning, 10*, 263–271.

Barley, Z. A., & Jenness, M. (1993). Cluster evaluation: A method to strengthen evaluation in smaller programs with similar purposes. *Evaluation Practice*, 14, 141–147.

Brandon, P.R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation*, *19*(3), 325–337.

Brandon, P.R., & Singh, J. M. (2009). The strength of the methodological warrants for the findings of research on program evaluation use. *American Journal of Evaluation*, *30*, 123–157.

Burke, B. (1998). Evaluating for a change: Reflections on participatory methodology. *New Directions for Evaluation*, *80*, 43–56.

Christie, C. (2007). Reported influence of evaluation data on decision makers actions: An empirical examination. *American Journal of Evaluation*, *28*(1), 8–25.

Cousins, J. B. & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, *56*(3), 331–364.

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 245–268). Boston: Kluwer Academic Publishers.

Cousins, J. B., & Earl, L. M., Editors. (1995). Participatory evaluation in education: Studies in evaluation use and organizational learning. London: Falmer.

Cousins, J. B., & Shulha, L. M. (2006). A comparative analysis of evaluation utilization and its cognate fields of inquiry: Current trends and issues. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 266–291). Thousand Oaks, CA: Sage.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80*, 5–23.

Cousins, J.B., Donohue, J. J., & Bloom, G.A. (1996). Collaborative evaluation in North America: Evaluators' self-reported opinions, practices and consequences. *Evaluation Practice*, *17*(3), 207–226.

Dillman, D. A. (2007). Mail and internet surveys: The tailored design method (Third ed.). New York: John Wiley & Sons, Inc.

Fetterman, D. M. (1994) Empowerment evaluation. *Evaluation Practice*, *15*(11), 1–15.

Gold, N. (1983). Stakeholder and program evaluation: Characterizations and reflections. *New Directions for Evaluation*, *17*, 63–72.

Greene, J. C. (1987). Stakeholder participation in evaluation design: Is it worth the effort? *Evaluation and Program Planning*, *10*(4), 379–394.

Greene, J. C. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, *12*(2*)*, 91–116.

Greene, J.C. (2000). Challenges in practicing deliberative democratic evaluation. *New Directions for Evaluation, 85*, 13–26.

Greene, J.C. (2005). A value-engaged approach for evaluating the Bunch-Da Vinci Learning Academy. *New Directions for Evaluation*, *106*, 27–45.

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Thousand Oaks, CA: Sage.

Henry, G. T. & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*. *24*(3), 293–314.

House, E. R. & Howe, E. K. (2000). Democratic deliberative evaluation. *New Directions for Evaluatio*n, *85*, 3–12.

House, E. R. (1980). *Evaluating with validity.* Beverly Hills, CA: Sage.

House, E.R. (2003). Stakeholder bias. *New Directions for Evaluation*, *97*, 53–56.

Johnson, K. A., Greenseid, L., Toal, S.A., King, J. A., Lawrenz, F. & Volkov, B. (2009). Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, *30*(3), 377–410.

Johnson, K.A (2008). Developing and validating scales of evaluation use in multi-site STEM education evaluations. *American Evaluation Association Conference*. Denver, CO.

Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, *21*(1), 93–110.

King, J. A. & Stevahn, L. (2002). Three frameworks for considering evaluator role. In K. E. Ryan and T. A. Schwandt (Eds.), Exploring evaluator role and identity. Greenwich, CT: Information Age Publishing.

King, J. A. (1988). Research on evaluation use and its implications for evaluation research and practice. *Studies in Educational Evaluation*, *14*, 285–299.

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation*, *80*, 57–67.

King, J. A., & Ehlert, J.C. (2008). What we learned from three evaluations that involved stakeholders. *Studies in Educational Evaluation*, *34*, 194–200.

King, J. A., & Pechman, E. M. (1984). Pinning a wave to the shore: Conceptualizing evaluation use in school systems. *Educational Evaluation and Policy Analysis*, *6*(3), 241–251.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*. *88*, 5–23.

Lambert, Z., & Durand, R. (1975), Some Precautions in Using Canonical Analysis. *Journal of Marketing Research, 12* (November), 468–75.

Lawrenz, F., & Huffman, D. (2003). How can multi-site evaluation be participatory? *American Journal of Evaluation*, *24*(4), 331–338.

Lawrenz, F., & King, J. A. (2009) Report 8: Beyond evaluation use cross case report. Retrieved June 15, 2009from http://www.cehd.umn.edu/projects/beu/documents.html

Leff, H. S., & Mulkern, V. (2002). Lessons learned about science and participation from multisite evaluations. *New Directions for Evaluation*, *94*, 89–100.

Levin, B. (1993). Collaborative research in and with organizations. *Qualitative Studies in Education*, *6*(4), 331–340.

Leviton, L. C., & Hughes, E. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review*, *5*, 525–548.

Mark, M. M. & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*, 35–56.

Mark, M. M., & Shotland, L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review*, *9*(5), 605–626.

Mark, M. M., Henry, G. T., & Julnes, G. (1999). Toward an integrative framework for evaluation practice. *American Journal of Evaluation*, *20*(2), 177–198.

Mercier, C. (1997). Participation in stakeholder-based evaluation: A case study. *Evaluation and Program Planning*, *20*(4), 467–475.

Patton, M. Q. (2008). *Utilization-focused evaluation (4th edition).* Thousand Oaks, CA: Sage Publications.

Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D., & Blyth, D. A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. In C. Weiss (Ed.) *Using social science research in public policy making*. Lexington, MA: Lexington Books.

Patton, M.Q. (1978). *Utilization-focused evaluation*. Newbury Park, CA: Sage Publications.

Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. *Evaluation Practice*, *18*, 209–225.

Rich, R. F. (1977). Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding. In C. Weiss (Ed.) *Using social research in public policy making*. Lexington, MA: Lexington Books.

Roseland, D.L., Volkov, B., Johnson, K., Lawrenz, F., & King, J. A. (2009) Beyond evaluation use: Field survey report. Retrieved on June 15, 2009 from http://www.cehd.umn.edu/projects/beu/documents.html

Shadish, W.R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shulha, L. M. & Cousins, J. B. (1997). Evaluation use: theory, research, and practice since 1986. *Evaluation Practice*, *18*(3), 195–208.

Sinacore, J. M. & Turpin, R. S. (1991). Multiple sites in evaluation research: A survey of organizational and methodological issues. *New Directions for Evaluation*, *50*, 5–18.

Stake, R. (1999, April). Representing quality in evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED437422.)

Straw, R. B., & Herrell, J. M. (2002). A framework for understanding and improving multisite evaluation. *New Directions for Evaluation*, *94*, 5–15.

Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, *15* (3), 321–338.

Taut, S. (2008). What have we learned about stakeholder involvement in program evaluation? *Studies in Educational Evaluation*, *34*, 224–230.

Themessl-Huber, M.T., & Grutsch, M.A. (2003). The shifting locus of control in participatory evaluations. *Evaluation*, *9*(1), 92–111.

Toal, S. A. (2007). *The development and validation of an Evaluation Involvement Scale for use in multi-site evaluation* (Unpublished doctoral dissertation). University of Minnesota, Twin Cities.

Toal, S. A. (2009). The validation of the evaluation involvement scale for use in multisite settings. *American Journal of Evaluation*, *30*(3), 349–362.

Toal, S. A., King, J. A., Johnson, K., & Lawrenz, F. (2009). The unique character of involvement in multi-site evaluation settings. *Evaluation and Program Planning*, *32*, 91–98.

Torres, R.T., Padilla Stone, S., Butkus, D.L., Hook, B. B., Casey, J., & Arens, S.A. (2000). Dialogue and reflection in a collaborative evaluation: Stakeholder and evaluator voices. *New Directions for Evaluation*, *85*, 27–38.

Turnbull, B. (1999). The mediating effect of participation efficacy on evaluation use. *Evaluation and Program Planning, 22,* 131–140.

Wallace, T. L. (2008). Integrating participatory elements into an effectiveness evaluation. *Studies in Educational Evaluation*, *34*, 201–207.

Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. In C.H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education.* (pp. 318–326). Boston, MA: Allyn and Bacon, Inc.

Weiss, C. H. (1980). *Social science research and decision making*. New York: Columbia University Press.

Weiss, C. H. (1990). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, *9*(1), 5–19.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, *19*(1), 21–33.

Worthen, B. R., & Schmitz, C. C. (1997). Conceptual challenges confronting cluster evaluation. *Evaluation*, *3*, 300–310.

APPENDIX A

NSF PROGRAM AND PROGRAM EVALUATION DESCRIPTIONS

The following tables describe the four National Science Foundation programs that were studied as a part of this research and that formed the basis of the Beyond Evaluation Use Research project at the University of Minnesota.

ADVANCED TECHNOLOGICAL EDUCATION (ATE) PROGRAM	
ATE PROGRAM DESCRIPTION	
	Program Period: 1993-2005 **Target:** Primarily students and teachers in 2-year colleges and collaborating institutions. Also secondary students and teachers, as well as 4-year post-secondary. **Sites Funded**: 345 sites were funded, of which 200 were 2-year colleges; secondary schools, 4-year colleges, and associations. **Program Budget:** Approximately $350 million **Program Purpose:** To increase U.S. productivity and international competitiveness by (1) building capacity to provide advanced technological education in high technology fields, and (2) increasing the number and skill of advanced technicians. Aimed primarily at 2-year colleges, but also included 4-yearcolleges and secondary schools.
ATE CORE EVALUATION DESCRIPTION	
ATE CORE EVALUATION QUICK FACTS	**Period:** 1999–2005 **Budget:** $3.1 million **Principal Investigator:** Arlen Gullickson, PhD., Professor Emeritus, Evaluation Center at Western Michigan University
CORE EVALUATION PURPOSE	The purpose of the ATE core evaluation was to measure the activities, accomplishments, and effectiveness of ATE projects and centers for general accountability purposes. To collect data on the underlying drivers of program success including collaboration with partners, professional development, and project sustainability.
CORE EVALUATION DESIGN	The evaluation design included mixed methods and featured a quantitative annual web-based survey and multiple site visits where primarily qualitative information was gathered. The surveys collected annual data on the activities and accomplishments and site visits collected data on collaboration with partners, professional development, and project sustainability.
CORE EVALUATION QUESTIONS	• *To what degree is the program achieving its goals?* • *Is it making an impact and reaching the individuals and groups intended?* • *How effective is it when it reaches its constituents?* • *Are there ways the program can be significantly improved?*
CORE EVALUATION ACTIVITIES	Principal evaluation activities included: • Convened meetings of ATE Advisory Committee • Fielded annual grantee survey 2000-2005 • Conducted 13 site visits—reports given only to sites • Commissioned 9 issue papers—synthesizing site visit findings and survey results • Developed four targeted studies (on value-added to business/industry, materials development, professional development, and sustainability.) • Four meta-evaluations were conducted.

COLLABORATIVES FOR EXCELLENCE IN TEACHER PREPARATION (CETP) PROGRAM	
CETP PROGRAM DESCRIPTION	
	Program Period: 1993–2000 (last year new projects were funded) **Target Population:** Prospective preK–12 teachers **Program Budget:** $350 million **Site Funds:** 25 Response to the national need to produce and retain increasing numbers of well qualified teachers of mathematics and science. **Program Purpose**: To achieve significant and systemic improvement in the science, technology, engineering, and mathematics (STEM) preparation of prospective pre-Kindergarten through grade 12 (preK–12) teachers.
CETP CORE EVALUATION DESCRIPTION	
CORE EVALUATION QUICK FACTS	**Period:** 1999–2004 **Budget:** $999,000 **Principal Investigator:** Frances Lawrenz, Ph.D., Associate Vice President for Research and Professor of Educational Psychology, University of Minnesota
CORE EVALUATION PURPOSE	The purpose of the CETP core evaluation was to learn to what extent the CETPs succeeded in achieving significant and systemic improvement in the science, technology, engineering, and mathematics (STEM) preparation of prospective pre-Kindergarten through grade 12 (preK-12) teachers.
CORE EVALUATION DESIGN	The overall design was mixed methods. Methods used included surveys (dean/department chair survey, PI/evaluator survey, pre- and post-faculty survey, college student survey, grades 6 to 12 student survey, K–12 teacher survey, principal survey, NSF scholars' surveys), classroom activities assessment rubric, and classroom observations. Although standardized instruments were developed, sites were free to use their own evaluation instruments or they could add items to the standard instrument. As the sites were not required to participate in the evaluation, data were not collected from all sites.
CORE EVALUATION QUESTIONS	• *To what extent did the CETP program impact the collaboration and focus of university faculty on instructional issues?* • *To what extent did the CETP program impact the instructional techniques used by university faculty?* • *Did K–12 teachers who participated in CETP projects view their preparation programs differently from teachers who participated in other preparation programs?* • *Were the instructional practices exhibited by K–12 teachers who participated in CETP projects different from the instructional practices exhibited by teachers who participated in other preparation programs?*
CORE EVALUATION ACTIVITIES	• Convened meetings with CETP project personnel. • Developed data collection instruments (surveys, classroom observation protocols.) • Provided technical assistance to local CETPs for data collection and analysis. • Standardized instruments were developed, but sites were free to use their own evaluation instruments or could add items to the standard instrument. • As the sites were not required to participate in the evaluation, data were not collected from all sites.

LOCAL SYSTEMIC CHANGE (LSC) THROUGH TEACHER ENHANCEMENT PROGRAM	
LSC PROGRAM DESCRIPTION	
	Program Period: 1995–2005 **Program Budget:** $250 million in the 10-year period **Target Population:** K–12 teachers of science and mathematics; focus on entire school systems or districts, not on individual teachers **Sites Funded**: 88 projects current and completed projects across 31 states, and involving 70,000 teachers, 4,000 schools, 467 school districts, and 2,142,000 students. **Program Purpose**: To enhance teachers' content and pedagogical knowledge and their capacity to use instructional materials. LSC requires that the professional development is delivered to all teachers in a system (building- or district-wide) not to individual teachers. The ultimate goal is improved student achievement in math and science.
LSC CORE EVALUATION DESCRIPTION	
CORE EVALUATION QUICK FACTS	**Period**: 1995–2005 **Budget**: $6.25 million **Principal Investigator**: Iris Weiss, Ph.D., President, Horizon Research, Inc.
CORE EVALUATION PURPOSE	The purpose of the LSC evaluation was two-fold: 1. To provide information that could be aggregated across projects, to enable NSF to report on progress to Congress and to make mid-course adjustments to the program; and 2. To assess individual projects and to provide for appropriate mid-course adjustments.
CORE EVALUATION DESIGN	A cross-project "core" evaluation system used a mixed-methods design, with mandatory participation by 88 local projects nationwide. The core evaluation allowed local evaluators to assess their own projects, but also allowed for aggregate data across projects yielding broader insights about the design, quality, and impact of the program as a whole. Project evaluators collected data using standardized questionnaires and interviews, as well as observation protocols designed to answer core evaluation questions. Evaluators completed ratings on the quality of LSC professional development programs.
CORE EVALUATION QUESTIONS	• *What is the overall quality of the LSC professional development activities?* • *What is the extent of school and teacher involvement in LSC activities?* • *What is the impact of the LSC professional development on teacher preparedness, attitudes, and beliefs about science and mathematics teaching and learning?* • *What is the impact of the LSC professional development on classroom practices in science and mathematics?* • *To what extent are the school and district contexts becoming more supportive of the LSC vision for exemplary science and mathematics education?* • *What is the extent of institutionalization of LSC reforms?*
CORE EVALUATION ACTIVITIES	Overall, the LSC core evaluation logged observations of 2,400 professional development sessions and 1,620 mathematics or science lessons, the completion of 75,000 teacher questionnaires, 17,380 principal questionnaires, and 1,782 teacher interviews.

MATH & SCIENCE PARTNERSHIPS—RESEARCH, EVALUATION, AND TECHNICAL ASSISTANCE (MSP-RETA) PROGRAM	
MSP PROGRAM DESCRIPTION	
	Program Period: October 2002–present **Program Budget**: ~ $460 million **Target Population**: Math and Science Educators K–12 and higher education **Sites Funded**: 77 projects in 30 states plus Puerto Rico, 550 school districts, 3300 schools, over 150 higher education institutions, and over 70 business partners. **Program Purpose**: Seeks to improve student outcomes in mathematics and science for all students, at all K–12 levels, and to significantly reduce achievement gaps in the mathematics and science performance of diverse student populations.
MSP-RETA DESCRIPTION	
EVALUATION QUICK FACTS	**Period**: October 2002 - 2007 **Budget**: $1.8 million **Principal Investigator**: Catherine Callow-Heusser, Project Director and Principal Investigator, Utah State University
EVALUATION PURPOSE	The purpose of the Utah State MSP-RETA was to provide technical assistance to MSP projects to identify evaluation needs and develop better program evaluation models.
EVALUATION ACTIVITIES	Provide curriculum development, professional development, career pathways and applied research on technician education through four project types: 1. Comprehensive Partnerships that implement change across the K-12 continuum in mathematics and/or science; 2. Targeted Partnerships that focus on a narrower grade range or disciplinary focus in mathematics and/or science; 3. Institute Partnerships: Teacher Institutes for the 21st Century that support the development of school-based teacher intellectual leaders; and 4. Research, Evaluation and Technical Assistance (RETA) projects that develop tools to assess the partnerships' progress and make their work more strategic, build evaluation capacity and conduct focused research.

APPENDIX B

SURVEY INVITATION AND REMINDER EMAILS

SURVEY INVITATION: VERSION 1 (LSC)

Dear XXXX,
The National Science Foundation (NSF) wants to improve the use of program evaluation. As a recent participant in a Local Systemic Change Initiative (LSC) grant, you are uniquely qualified to provide essential and firsthand input about your experiences with the LSC program evaluation. The survey, accessible through the link listed below, asks for your candid perceptions of your experience, which will help us better understand the evaluation process and its outcomes. This information will, in turn, enable NSF to enhance its evaluation activities. The web-based survey takes around 10 minutes to complete.

No names will be used in this study, and only aggregated results will be reported. At the end of the survey, you will be asked for permission to contact you if you are selected for a follow-up interview.

To show our appreciation for your time and thoughts, the first 50 respondents will receive a $10 amazon.com gift certificate. In addition, everyone who responds to the survey will be entered in a drawing for one of several web cameras. A webcam will add a visual dimension to your online chatting, videoconferencing, and e-mailing, allowing you to enjoy videoconferencing with friends and family all over the world with high-quality audio and video.

To participate, click on the survey link below, and you will be prompted for your pre-assigned account name and password (provided at the end of this email).

Please feel free to contact us at the email addresses listed below if you have questions or concerns about this survey. Or, if would like to talk to someone else, you may contact the University of Minnesota's Research Subjects' Advocate Line at (612) 625-1650.

We recognize the many demands made on your time, and thank you in advance for your willingness to help. Your input is critical, and we appreciate your involvement.

Sincerely,
Frances Lawrenz, Professor
Educational Psychology—University of Minnesota
lawrenz@umn.edu

Jean A. King, Professor
Educational Policy & Administration—University of Minnesota
kingx004@umn.edu

Link: http://www2.education.umn.edu/EdPsy/NSF/Survey/Default.asp

Your account name:
Your password:
This study is sponsored by NSF number REC0438545

SURVEY INVITATION: VERSION 2 (MSP, CETP, ATE)

Dear XXX,
The National Science Foundation (NSF) wants to improve the use of program evaluation. As a recent participant in a Math Science Partnership (MSP) project, you are uniquely qualified to provide essential and firsthand input about your experiences with the Utah State Math Science Partnership-Research Evaluation Technical Assistance (MSP-RETA) evaluation project. This study is about the MSP-RETA evaluation activities conducted by the Utah State University team; it is not about your project-level evaluations. The survey, accessible through the link listed below, asks for your candid perceptions of your experience, which will help us better understand the evaluation process and its outcomes. This information will, in turn, enable NSF to enhance its evaluation activities. The web-based survey takes around 10 minutes to complete.

No names will be used in this study, and only aggregated results will be reported. At the end of the survey, you will be asked for permission to contact you if you are selected for a follow-up interview.

To show our appreciation for your time and thoughts, the first 50 respondents will receive a $10 amazon.com gift certificate. In addition, everyone who responds to the survey will be entered in a drawing for one of several web cameras. A webcam will add a visual dimension to your online chatting, videoconferencing, and e-mailing, allowing you to enjoy videoconferencing with friends and family all over the world with high-quality audio and video.

To participate, click on the survey link below, and you will be prompted for your pre-assigned account name and password (provided at the end of this email).

Please feel free to contact us at the email addresses listed below if you have questions or concerns about this survey. Or, if would like to talk to someone else, you may contact the University of Minnesota's Research Subjects' Advocate Line at (612) 625-1650.

We recognize the many demands made on your time, and thank you in advance for your willingness to help. Your input is critical, and we appreciate your involvement.

Sincerely,
Frances Lawrenz, Professor
Educational Psychology, University of Minnesota
lawrenz@umn.edu

Jean A. King, Professor
Educational Policy & Administration, University of Minnesota
kingx004@umn.edu

Link: http://www2.education.umn.edu/EdPsy/NSF/Survey/Default.asp

Your account name:
Your password:
This study is sponsored by NSF number REC0438545

REMINDER EMAIL: VERSION 1 (LSC)

Dear XXX,
You have recently received an e-mailed invitation to complete a survey funded by the National Science Foundation (NSF). In an effort to improve the use of program evaluation, we are surveying Local Systemic Change (LSC) grant recipients because we feel you are uniquely qualified to provide essential and firsthand input about your experiences with the CETP program evaluation. The linked survey asks for your candid perceptions of your experience, which will help us better understand the evaluation process and its outcomes. This information will, in turn, enable NSF to enhance its evaluation activities. The web-based survey takes around 10 minutes to complete.

No names will be used in this study, and only aggregated results will be reported. At the end of the survey, you will be asked for permission to contact you if you are selected for a follow-up interview.

To show our appreciation, everyone who responds to the survey will be entered in a drawing for one of several web cameras. A webcam will add a visual dimension to your online chatting, videoconferencing, and e-mailing, allowing you to enjoy videoconferencing with friends and family all over the world with high-quality audio and video.

To participate, click on the survey link below, and you will be prompted for your pre-assigned account name and password (provided at the end of this email). Please feel free to contact us at the email addresses listed below if you have questions or concerns about this survey. Or, if would like to talk to someone else, you may contact the University of Minnesota's Research Subjects' Advocate Line at (612) 625-1650.

We recognize the many demands made on your time, and thank you in advance for your willingness to help. Your input is critical, and we appreciate your involvement.

Sincerely,
Frances Lawrenz
Professor, Educational Psychology
University of Minnesota
lawrenz@umn.edu

Jean A. King
Professor, Educational Policy & Administration
University of Minnesota
kingx004@umn.edu

Link: http://www2.education.umn.edu/EdPsy/NSF/Survey/Default.asp

Your account name:
Your password:
This study is sponsored by NSF number REC0438545

REMINDER EMAIL: VERSION 2 (MSP, CETP, ATE)

Dear XXX,
Recently, we sent you a survey from a project funded by the National Science Foundation (NSF). If you have already completed this survey, thank you for your participation. If you haven't had a chance to complete the survey yet, we hope you will take 5 minutes out of your busy day to visit our survey site at your earliest opportunity. Your personalized URL and login information for this survey can be found at the bottom of this message.

As a participant in this research, you are making an important contribution and helping shape the future of our field. Thank you.

Frances Lawrenz
Professor, Educational Psychology
University of Minnesota
lawrenz@umn.edu

Jean A. King
Professor, Educational Policy & Administration
University of Minnesota
kingx004@umn.edu

To participate, click on the survey link below, and you will be prompted for your preassigned account name and password.

Link: http://www2.education.umn.edu/EdPsy/NSF/Survey/Default.asp

Your account name:
Your password:
This study is sponsored by NSF number REC0438545

APPENDIX C

INTERVIEW CONSENT FORM

CONSENT FORM
Evaluation Use Study

You are invited to be in a research study of the use and influence of program evaluation. You were selected as a possible participant because you have participated in the evaluation or are aware of the results. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Frances Lawrenz and Jean King at the University of Minnesota with the assistance of graduate students Lija Greenseid, Kelli Johnson, Stacie Toal and Boris Volkov.

Background Information

The purpose of this study is: To investigate the possible influence of the results of program evaluations.

Procedures

If you agree to be in this study, we would ask you to do the following things:
Answer questions about the use of evaluation results from the [**NAME]** evaluation. Your answers to these questions will be recorded and transcribed but your name will not be directly associated with comments in any reports. The interview should take 30-60 minutes. You may also be asked to provide us with any artifacts demonstrating the influence of the evaluation on your project.

Risks and Benefits of being in the Study

The study has the risk that thinking about the evaluation results may bring up old memories or frustrations. You will also be giving up some of your time.

The benefits to participation are that you will be helping us to understand the extent and type of influence of program evaluations so that they may be better designed in the future.

Confidentiality

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. Tapes will be stored in a locked file and erased after transcription. Only the researcher and the transcriber will have access to the original tape.

Voluntary Nature of the Study

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota or the National Science Foundation. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions

The researchers conducting this study are: Frances Lawrenz and Jean King. If you have any questions, you are encouraged to contact them at University of Minnesota, 612-625-2046, lawrenz@umn.edu or kingx004@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), you are encouraged to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

Please print a copy of this information to keep for your records.

APPENDIX D

INTERVIEW PROTOCOL

(GENERIC, NOT TAILORED TO EACH PROGRAM)

Begin with IRB information. Get their consent to be interviewed recorded on the tape.

Introduction
Thank you for agreeing to talk with us about your experiences with the ATE program evaluation. First, so we can have it on the tape, would you please introduce yourself by stating your name and your relationship to your ATE project at the time of the program evaluation? And how long were you involved with the ATE program evaluation?

Clarifying Questions
First, we want to clarify that we will be asking questions about your experiences with the XXX program evaluation that took place between XXX. You may remember there were two evaluations associated with your grant: an individual project evaluation that you or your project was responsible for, and the broader program evaluation conducted XXX.

As part of the program evaluation you may remember the {items specific to program evaluation)

1. Is it clear that we want to focus on the program evaluation and that you understand what the program evaluation is?

2. By the way—do you remember taking a survey from us about this same topic last fall? (If yes) Thinking back on the survey, did you understand then that we were asking questions about the program evaluation?

Program Evaluation Involvement
We have two main areas to discuss today: involvement and impact. Let's begin by having you describe the ways in which you and your project were involved in the program evaluation activities.

3. Could you please describe how you and your project were involved with program evaluation activities such as requesting the evaluation, designing the parts of the program evaluation, collecting data, or analyzing and reporting findings?

   Probes, if necessary:
   - Were you involved in the planning stages of the evaluation?
     - Did you provide input on evaluation questions the program evaluation should address?
   - Did you help in developing or field testing the survey instruments?
   - Did your project provide data for the annual survey?
   - Did you help create the site visit protocol?
   - Did you review or draft any reports of the program evaluations findings?

4. What motivated you to get involved in these ways?

5. To what extent do you feel the input and/or assistance you provided was used, or wasn't used, by the program evaluation staff?

6. Overall, how would you rate your level of involvement in the program evaluation? Would you say you were not involved, involved only a little, involved some, or involved extensively?

Impact

Now we're going to ask you some questions about the effect the program evaluation had on your project, followed by questions about the impact on you, and then about broader impacts.

Project Level Impact

7. First, in what ways, if any, did the program evaluation have an impact on your project?

   Probes:
   a. Did you use the data provided by the evaluation in your project's evaluation?
   b. Did you use any of the findings from the program evaluation to make decisions about your project? If so, which ones?
   c. Did the program evaluation findings spark any conversations within your project or among your project's stakeholders?
   d. Did you use the findings to advocate for support for your project?
   e. Did you take any of the action steps recommended in the targeted brochures, for example, sustainability, material development, professional development, etc?
   f. Did the program evaluation process affect the way the project staff interacted?
   g. Have you used the site visit protocols or the planning guide?
   h. Were there other ways in which you used the program evaluation findings?

8. Overall, how would you rate the level of impact the program evaluation had on your project? Would you say there was no impact, only a little impact, some impact, or an extensive amount of impact?

Individual Level Impact

Next let's focus more closely on the effect that the program evaluation process had on you (OR: each of you) personally. In what ways, if any, did the program evaluation process have an impact on you?

Probes:
   a. Did you learn new evaluation skills? (Probes, specific to program.)
   b. Did you learn any new knowledge about your project as a result of the evaluation?
   c. Did you feel differently about the importance of evaluation?

9. Overall, how would you rate the level of impact the program evaluation had on you? Would you say there was no impact, only a little impact, some impact, or an extensive amount of impact?

Factors Affecting the Program Evaluation's Impact

10. Ok, so you've described the different ways in which the program evaluation had an impact on you and your project. What do you think were the reasons that the study had the impact it did?

    Probes:
    a. Anything else that inhibited the impact the program evaluation had on you and your project?
    b. Anything else that enhanced the impact the program evaluation had on you and your project?

Broader Influence of the Program Evaluation

Now that we have talked about the impact that the program evaluation had on you and your project we'd like to look at the effects of the program evaluation more broadly.

11. First, however, what did you think was the goal of the program evaluation?

12. Can you think of any ways in which the program evaluation had an effect on the STEM education field?

13. What about any effects of the program evaluation on the evaluation community?

14. Were there any other broad impacts of the program evaluation, from your perspective? If no—do you think it's realistic to expect broad impacts from these types of large-scale program evaluations?

15. What do you believe it would have taken for the program evaluation to have had a greater impact on the field?

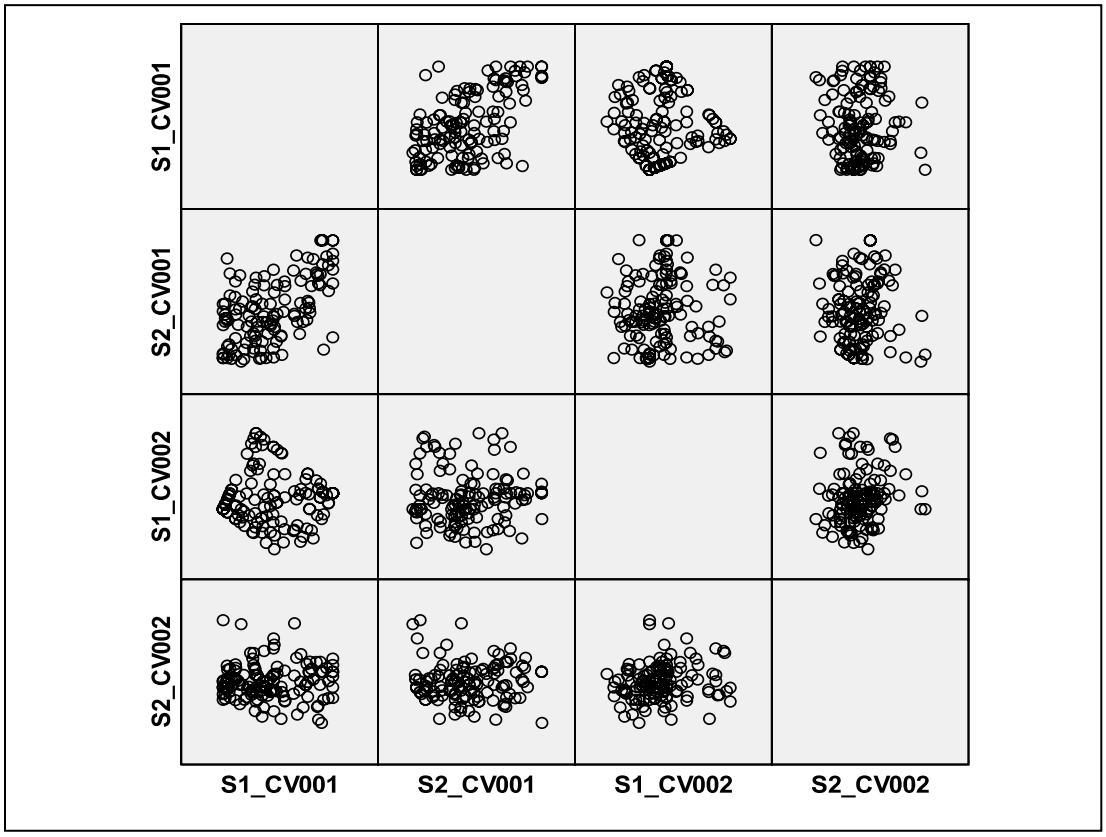
Wrapping-up

16. In sum, what do you feel is the most important lasting effect of the program evaluation, if any?

17. Are there any other issues regarding the impact of the program evaluation that you think are important for us to hear?

APPENDIX E

BIVARIATE SCATTERPLOTS PRODUCED TO CHECK STATISTICAL ASSUMPTIONS OF CCA



In the figure above, we see 16 scatterplots combined into a single figure that show the combination of the factors scores of respondents related to the two factor sets and the canonical functions created to explain the nature of the relationship between involvement in and use of evaluation. When examining the figure, S1 represents the involvement factor scores and S2 represents the use/influence factor scores in relationship to CV001 (the canonical function score 1) and CV002 (the canonical function score 2). The data

appears to meet all assumptions because there does not appear to be a pattern in the scatterplots, and there are not large differences in how spread out each scatterplot appears.